150.8 A673 no.163 c.1
Woodworth, Robert Sessions,
Archives of psychology. --
R.W.B. JACKSON LIBRARY
OISE CIR
3 0005 03037 6274

SELF-ESTIMATES OF IMPROVEMENT IN REPEATED TASKS

BY
NATALIE KNEELAND, Ph.D.

ARCHIVES OF PSYCHOLOGY
R. S. WOODWORTH, EDITOR
No. 163

150.8
A673
no. 163

NEW YORK
March, 1934

LIBRARY
FEB 21 1973
TH[illegible]
FOR STUDIES IN EDUCATION

DEPARTMENT OF PSYCHOLOGY
LIBRARY
UNIVERSITY OF PORTLAND

ARCHIVES OF PSYCHOLOGY

COLUMBIA UNIVERSITY, NEW YORK CITY

The Subscription price is six dollars per volume of about 500 pages. Volume I comprises Nos. 2–10; Volume II, Nos. 11–18; Volume III, Nos. 19–25; Volume IV, Nos. 26–32; Volume V, Nos. 33–39; Volume VI, Nos. 40–46; Volume VII, Nos. 47–52; Volume VIII, Nos. 53–58; Volume IX, Nos. 59–63; Volume X, Nos. 64–68; Volume XI, Nos. 69–73; Volume XII, Nos. 74–78; Volume XIII, Nos. 79–85; Volume XIV, Nos. 86–91; Volume XV, Nos. 92–98; Volume XVI, Nos. 99–104; Volume XVII, Nos. 105–112; Volume XVIII, Nos. 113–120; Volume XIX, Nos. 121–127; Volume XX, Nos. 128–133; Volume XXI, Nos. 134–139; Volume XXII, Nos. 140–146; Volume XXIII, Nos. 147–154. The available numbers are as follows:

7. Statistical Study of Literary Merit: F. L. Wells. 30c.
8. Magnitude of Stimulus and Time of Reaction: Sven Froeberg. 35c.
9. Perceptual Factors in Reading: F. M. Hamilton. 50c.
10. Time in English Verse Rhythm: W. Brown. 70c.
11. Hearing of Primitive Peoples: F. G. Bruner. $1.00.
12. Studies in Development and Learning: E. A. Kirkpatrick. $1.00
13. Inaccuracy of Movement: H. L. Hollingworth. 80c.
14. Quantitative Study of Rhythm: H. Woodrow. 60c.
15. Psychology of Efficiency: H. A. Ruger. $1.25.
16. Electrical Processes in the Human Body and their Relation to Emotional Reactions: F. L. Wells and A. Forbes. 40c.
17. Relative Merit of Advertisements: E. K. Strong, Jr. $1.00.
18. Attention and Movement in Reaction Time: J. V. Breitweiser. 50c. (Cl., 75c.)
19. Empirical Study of Certain Tests for Individual differences: M. T. Whitley. $1.25. (Cl., $1.50.)
20. Visual Acuity with Lights of Different Colors and Intensities: D. E. Rice. 50c. (Cl., 75c.)
21. Curve of Forgetting: C. H. Bean. 45c. (Cl., 75c.)
23. Reaction Time to Retinal Stimulation: A. T. Poffenberger, Jr. 70c. (Cl., 95c.)
24. Interference and Adaptability: A. J. Culler. 75c. (Cl., $1.00.)
25. Reaction to Multiple Stimuli: J. W. Todd. 60c. (Cl., 85c.)
26. Study in Incidental Memory: G. C. Myers. $1.00. (Cl., $1.25.)
27. Statistical Study of Eminent Women: C. S. Castle. 80c. (Cl., $1.05.)
28. Mental Capacity of the American Negro: M. J. Mayo. 60c. (Cl., 85c.)
29. Experimental Studies in Judgment: H. L. Hollingworth. $1.25. (Cl., $1.50.)
30. Psychological Researches of James McKeen Cattell: A Review by Some of His Pupils. $1.00.
31. Fatigue and Its Effects upon Control: I. E. Ash. 60c. (Cl., 85c.)
32. Transfer of Practice in Cancellation: M. A. Martin. 60c. (Cl., 85c.)
33. Intellectual Status of Children Who are Public Charges: J. L. Stenquist, E. L. Thorndike and M. R. Trabue. 50c. (Cl., 75c.)
34. Quickness of Learning and Retentiveness: D. O. Lyon. 50c. (Cl., 75c.)
35. Overcoming of Distraction: J. J. B. Morgan. 75c. (Cl., $1.00.)
36. Psychology of the Negro: G. O. Ferguson, Jr. $1.25. (Cl., $1.50.)
37. Effect of Distraction on Reaction Time: J. E. Evans. $1.00. (Cl., $1.25.)
38. Effect of Humidity on Nervousness and on General Efficiency: L. I. Stecher. 90c. (Cl., $1.15.)
39. Mechanism of Controlled Association: M. A. May. 75c. (Cl., $1.00.)
41. Mental Fatigue during Continuous Exercise of a Single Function: T. R. Garth. 85c. (Cl., $1.10.)
42. Psychological Study of Trade-Mark Infringement: R. H. Paynter. 85c. (Cl., $1.10.)
43. Individual Differences and Family Resemblances in Animal Behavior: H. J. Bagg. 70c. (Cl., $1.00.)
44. Experiment Studies in Recall and Recognition: E. M. Achilles. 90c. (Cl., $1.25.)
45. Morphologic Aspect of Intelligence: Sante Naccarati. 70c.
47. Effects of Practice on Judgments of Absolute Pitch; E. Gough. $1.25.
48. Experimental Study of Silent Thinking: R. S. Clark. $1.40.
49. Some Empirical Tests in Vocational Selection: H. W. Rogers. 75c.
50. Adenoids and Diseased Tonsils: Their Effect on General Intelligence: M. Cobb. $1.00.
51. Experimental Study of the Factors and Types of Voluntary Choice: A. H. Martin. $1.50.
52. Some Well-known Mental Tests Evaluated and Compared: D. R. Morgenthau. 80c.
53. Mood in Relation to Performance: E. T. Sullivan. $1.00.
54. Influence of Incentive and Punishment upon Reaction Time: A. M. Johanson. 80c.
55. Psychological Tests Applied to Factory Workers: E. T. Burr. $1.25.
56. Study of the Relation of Accuracy to Speed: H. E. Garrett. $1.25.
57. Experimental Study of Hunger in Its Relation to Activity: T. Wada. $1.50.
58. Individual Differences as Affected by Practice: G. S. Gates. $1.00.
59. Studies in Industrial Psychology: E. O. Bregman. 90c.
60. The Mental Status of Psychoneurotics: A. D. Tendler. $1.25.
61. Effects of Attention on the Intensity of Cutaneous Pressure and on Visual Brightness: S. M. Newhall. $1.25.
62. Measurement of Motor Ability: E. G. Garfiel. 90c.
63. Race Differences in Inhibition: A. L. Crane. $1.50.
64. Individual Differences in Incidental Memory: S. M. Shellow. $1.25.
65. Character Traits as Factors in Intelligence Test Performance: W. M. Brown. $1.25.
66. Study of the Sexual Interest of Young Women: F. I. Davenport. $1.25.
67. Psychology of Confidence: W. C. Trow. $1.25.
68. Experimental Studies of College Teaching: H. E. Jones. $1.25.
69. Influence of Treatment for Intestinal Toxemia on Mental and Motor Efficiency: A. E. Paulsen. $1.00.
70. Study of Suggestibility of Children: M. Otis. $1.50.
71. Praise and Reproof as Incentives for Children: E. B. Hurlock. $1.00.
72. Experimental Study of Thinking: E. Heidbreder. $1.75.
74. Estimation of Time: R. Axel. $1.00.
77. Tested Mentality as Related to Success in Skilled Trade Training: T. M. Abel. $1.25.
78. Aggressive Behavior in a Small Social Group: E. M. Riddle. $1.75.
79. Memory Value of Advertisements: E. R. Brandt. $1.25.
80. Critical Examination of Test-Scoring Methods: R. G. Anderson. $1.00.
81. Thermal Discrimination and Weber's Law: E. A. K. Culler. $1.75.
82. Correlational Analysis of Typing Proficiency: L. Ackerson. $1.50.
83. Recall as a Function of Perceived Relations: C. B. Key. $1.25.
84. Consistency of Rate of Work: C. E. Dowd. $1.00.
85. Experimental Investigation of Recovery from Work: S. L. Crawley. $1.25.
86. Facilitation and Inhibition: T. N. Jenkins. $1.00.
87. Variability of Performance in the Curve of Work: J. D. Weinland. $1.00.

(*Continued on inside back cover.*)

SELF-ESTIMATES OF IMPROVEMENT IN REPEATED TASKS

BY
NATALIE KNEELAND, Ph.D.

ARCHIVES OF PSYCHOLOGY
R. S. WOODWORTH, Editor
No. 163

LIBRARY
FEB 21 1973
THE ONTARIO INSTITUTE
FOR STUDIES IN EDUCATION

NEW YORK
March, 1934

TABLE OF CONTENTS

PAGE

I. INTRODUCTION 5

II. AN EXPERIMENT IN SELF-ESTIMATION OF IMPROVEMENT IN THREE TASKS 9

III. PERFORMANCE IN THE THREE TASKS 16

IV. THE SUBJECTS' ESTIMATES 22

V. ESTIMATES AS RELATED TO IMPROVEMENT 32

VI. INDIVIDUAL CONSISTENCY IN ESTIMATES 46

VII. INTERPRETATION OF OUTSTANDING RESULTS 53

VIII. SUMMARY AND CONCLUSIONS 68

IX. BIBLIOGRAPHY 75

THE LIBRARY

The Ontario Institute

for Studies in Education

Toronto, Canada

Self-Estimates of Improvement in Repeated Tasks

I. INTRODUCTION

The process of self-estimation is found in all phases of activity. College students, on the basis of past experience, estimate beforehand what they may ''make'' on a quiz, and experience pleasure or chagrin according to their later estimates of actual performance. New sales people keep a watchful eye on their increase in sales during the early training period, while the old-timers compare their totals with last year's figures. Children are keen to notice improvement in school work or play.

Although this tendency to estimate is prevalent, we are struck by the fact that it is sometimes carried on in rather a hit or miss fashion. This point raises the question as to the validity of such estimates as actual measures of improvement (or of performance) and also the problem of the manner in which such estimates are made.

We would like to be able to answer such questions as: whether or not subjects can estimate their own improvement with any degree of accuracy; whether there is a tendency towards over- or underestimation on the part of some individuals; whether subjects are consistent in their estimates in different tasks, or even in the same task; and just how they go about estimating improvement. In order to get some light on the subject, let us first turn to previous work on self-estimates.

Summary of Work on Self-Estimates

Most of the work on self-estimates has been undertaken in connection with studies of assurance. Trow (22) studied the relationship between confidence in performance and accuracy of performance in a series of tasks, such as line and weight discrimination, judgment of hand-writing, memory span for digits, cancellation. Greene (13) utilized the answers to a variety of true-false tests, in which his subjects indicated the degree of confidence in their answers. Both of these investigators found that the degree of confidence in one's judgments of performance was not constant but varied with the nature of the situation. They also found little relationship between confidence and achievement. Filter (5) used a confidence scale in requiring subjects to estimate whether or not they could perform a series of complicated problems in a decoupagé and string manipulation test. He concluded that over-estimation could be taken as one measure of self-assurance.

Hollingsworth (16) in a study on judgments of improvement in terms of ''better than usual'' and ''worse than usual'' in tapping, color-naming and opposites found that an individual's judgment of efficiency or inefficiency possessed a degree of correctness that varied with the degree of his confidence. Other findings which have a bearing on the present study were the slight predisposition towards judgments of ''better'' and the fact that ''progressively larger variations in performance are required as the basis for judgments of a given degree of confidence as one passes from an automatic, objectively observable motor performance (such as tapping), through work involving perceptional reactions (color-naming), to work of a more strictly mental and less objectively observable character (opposites).'' In a somewhat similar experiment in which judgments were made of performance in color-naming, opposites, cancellation and addition, this same investigator found that the important criteria of judgments of efficiency in these tests were either directly perceptual in character or were inferences from perceptual data, and that affective processes did not play an important rôle.

Two studies on self-estimates that are not directly related to the problem of assurance are one by Hoffman (15) on self-estimates of personal character traits, and a study on estimates of the amount of work one can do, by Filter (6). Hoffman found that her subjects rated themselves higher in desirable than in undesirable traits, and discovered a negative relationship between the possession of a desirable trait and self-estimate on this trait. While this study has not a direct bearing on the estimation of self-improvement, the results are related in a general way to the problem of over- and under-estimation.

Filter had his subjects estimate the amount of work they could do in a given period of time. The six different tasks were naming the makes of automobiles, naming the capitals of states in the Union, giving the names of different magazines, recognizing nonsense syllables, working forty fractions, and adding forty columns of figures. In this study an actual estimate was called for, rather than a mere statement of assurance. The subjects had no previous knowledge of performance, and the estimates were made in terms of amount of work to be accomplished, not in terms of improvement to be made. Estimates after performance were not called for. Filter concluded from his study that there was no clearly demonstrable relationship between estimates and performance; that over- and under-estimations were largely determined by the nature of the task estimated; that the good performers showed a tendency to over-estimate and the poor performers to under-estimate; and finally, that there was little constancy of any one form of reaction in varying situations.

While these earlier studies afford valuable information on the subject of assurance as related to performance, they contribute little direct evidence on the estimation of the amount one can improve. Where improvement is brought in, it is expressed in the form of a judgment, such as ''better than usual,'' ''worse than usual,'' rather than as an estimate. However, in life situations it is frequently as important for an individual to estimate how much he will gain or lose as it is to simply judge whether he will, or will not be able to improve.

Definition of the Problem

The purpose of this investigation, therefore, is to study an individual's ability to estimate his own improvement, and to discover the nature of some of the factors at work during the estimating process.

Procedure for Studying Self-Estimates of Improvement

Since there was no available technique for measuring self-estimates of improvement, it was first necessary to devise a procedure that would secure as much information on the subject as possible in the shortest space of time, and still be flexible enough to be suited to various kinds of tasks presented by either the amount or time limit method, and performed either individually or in groups.

The plan, in brief, was to have a subject perform a task, score his first performance, and then repeat the task as many times as desired, estimating before and after each repetition the amount of improvement anticipated or actually made. At the end of the experimental period he was asked a variety of questions pertaining to his method of estimating.

To keep the data as objective as possible, the subject was required to estimate his improvement from an actual and known standard. In the experiment to be described, the score in the first trial, the *original* score, was used as the standard of reference. (If the subject had been allowed to score his performance after each trial, it would have been feasible to have used his score in one trial as the basis for estimates in the following trial, instead of having him go back to the first trial.)

In making his estimate, the subject was first requested to check one of three responses: *gain, loss,* or *same.* (Provision for a *same* response was made in order to study this particular type of estimate. Ordinarily, it would be omitted.) Checking the response in this manner helped to orient the subject in the estimating procedure.

Two types of estimates were requested: *think* and *may.* That is, the subject was asked to estimate first, how much he *thought* he would gain or lose, or had gained or lost; and then, how much he *might* gain or lose, or might have gained or lost. (When the subject checked a *same* response, he was told to indicate how much he might possibly gain and lose, or how much he might have possibly gained and lost.) The use of the terms *think* and *may* provided for an expression of upper and lower limits of estimation. This refinement in estimation is not essential to the general procedure, however, for it would be possible to omit the terms *think* and *may* altogether, and express the estimates simply in terms of future and past tense.

In order to impress the subject with the fact that he was to estimate performance in terms of improvement, estimates were expressed in increments of improvement over the standard, rather than in gross score, although either method would have given a measure of improvement. That is, the subject would say, ''I think I will improve 8 (over a score of 20),'' rather than, ''I think my score will be 28 this time.''

As a variation in estimating procedure, the subject was directed in one trial to assume the attitude that improvement would be made, and to set a goal of improvement, measured again from the standard. The after-estimates for this trial were made as usual.

II. AN EXPERIMENT IN SELF-ESTIMATION OF IMPROVEMENT IN THREE TASKS

In setting up an experiment in self-estimation of improvement the question arose as to whether it would be best to study the reactions of a number of individuals in a group, or to study a few individuals intensively. Because the investigator was mainly interested in discovering what occurred in everyday estimating situations, that is, with fairly naïve estimators rather than with thoroughly trained laboratory subjects, the group method of experimentation was decided upon. Furthermore, the group situation introduces a type of competition that frequently motivates self-estimation.

It was necessary to try out material and method of procedure with small groups first in order to discover what changes might need to be made. This experimental work was carried out during the year before the major research was conducted. As a result of this preliminary work it was decided to utilize student groups and to limit the experimental period to one class hour. Time limits for the tasks were set and the number of tasks reduced from four to three. Each of these tasks was to be performed once, scored by the subjects, and then repeated three times, making four trials in all, with requests for estimates before and after each repetition. The setting of a goal was introduced before the last trial.

A blank was prepared for the recording of estimates. (See Figure 1.) Spaces were left at the right-hand side for recording scores and improvement. The blanks were put up in three different colors, one for each task. The fact that this blank provided for the recording of all objective, numerical data made the statistical handling of the results a straight-forward one. The figures on the estimate blanks could be used directly in making all necessary computations.

Selection of Tasks

Cancellation, scribbling and knot tying were the three tasks selected as meeting the requirements of the experiment that (1) they could be easily and objectively scored; (2) they would provide for the possibility of improvement, with some variation, and that this improvement could be measured in well-defined units; (3) that they would be short and easily comprehended; and (4) that they

Original Score

Name..
Male
Female

Date..

SECOND TRIAL

Shall improve..
Shall not do so well..
Score will remain the same..

Before	*After*
Think I shall gain at least..............	Think I gained at least..............
May possibly gain as much as..............	May possibly have gained as much as..............
Think I shall lose at least..............	Think I lost at least..............
May possibly lose as much as..............	May possibly have lost as much as..............

THIRD TRIAL

Shall improve..
Shall not do so well..
Score will remain the same..

Before	*After*
Think I shall gain at least..............	Think I gained at least..............
May possibly gain as much as..............	May possibly have gained as much as..............
Think I shall lose at least..............	Think I lost at least..............
May possibly lose as much as..............	May possibly have lost as much as..............

FOURTH TRIAL

Before	*After*
Set as my goal an improvement of..............	Think I gained at least..............
	May possibly have gained as much as..............
	Think I lost at least..............
	May possibly have lost as much as..............

FIGURE 1. Estimate Blank

would be sufficiently different in character to avoid confusion of estimate.

The first and second halves of the Woodworth-Wells cancellation number group test comprised the material of the first task. The first half was used in the first and third trials, the second half in the second and fourth trials. The time limit of 35 seconds was set for each trial. Each number containing both a 6 and a 9 was to be cancelled.

In the scribbling task, the subject made a series of side-to-side oscillations with the pencil while drawing it down the page—a performance akin to tapping. The time allowed per trial was 5 seconds; two trials were made in immediate succession, and the better of the two was considered. The score was the number of bends at the right side of the scribble.

The knot-tying task consisted in tying knots as quickly as possible in a piece of twine a yard in length. This task involved a certain amount of skill and ingenuity in handling the material. The time limit was 1 minute. Twine of a different color but the same consistency was used for each trial.

SELECTION OF SUBJECTS

The basis of the selection of subjects for the experiment was one of availability. Through the courtesy of Dr. Schoen and Dr. Cleeton of the Carnegie Institute of Technology, opportunity was given to test students from the Arts school, the Science school, the school of Industries and the Margaret Morrison College for Women. Dr. Filter of the University of Pittsburgh put his class in general psychology at the writer's disposal. In order to secure a somewhat different group, permission was obtained from Miss Nina Clover to utilize the class hour of a group of boys and girls taking a course in retail selling in the South Hills High School, Pittsburgh. Dr. Greene of the Research Bureau for Retail Training made it possible to utilize the services of a small group of department store people who were enrolled in an executive training course.

A group of fifteen graduate students enrolled in the Research Bureau for Retail Training and a group of ten adults acted as subjects in two supplementary experiments.

The subjects in the main group were distributed according to classes as follows:

College groups—	
General psychology (Pitt.)	152 (2 sections)
Science (Sc.)	72
Arts	64
Industries (Ind.)	38

Margaret Morrison College (M.M.)	20
High School Group (H.S.)	42
Store Group (Executive training course)	18
Total	406

ADMINISTRATION

An effort was made to have all the envelopes containing the material distributed before the class hour was started. Those who arrived early were instructed not to open the envelopes. A supply of sharpened pencils was kept on hand. A seat was left vacant between each two subjects.

The experimenter began with two explanatory statements: first, that the tasks that would be given were not intelligence tests and that everyone would find them very simple to do; and second, that we were interested in their own individual scores and not concerned with comparing their scores with those of other members of the group.

General directions then followed. They included instructions in regard to seating arrangements, following directions, and opening envelopes. The seating arrangement was worked out in order to facilitate the scoring of the tasks after the first trial. Demonstration was used wherever possible. Throughout the directions that follow an asterisk before a statement will indicate that a demonstration of material or method was made.

The directions for each task were simple. They are given in outline form below:

DIRECTIONS FOR CANCELLATION

*Out of envelope 1 take the cancellation sheet with figures on it like this. Leave the rest of the material in the envelope.
Turn the cancellation sheet face down.
When I say ''Ready'' turn the sheet over right side up.
At the word ''Go'' start in the upper left-hand corner of the sheet and underline every series of figures containing both a 6 and a 9.
*Work across the sheet as in reading, starting at the left side for every line.
At the word ''Stop,'' stop immediately and make a circle around the series you were looking at when the command to stop was given.
Remember that only the series containing both a 6 and a 9 are to be underlined, and that you are to work across the page.
Accuracy will count as well as speed.
Ready—turn the sheet over.
Go! (Time limit, 35 seconds.)
Stop! Be sure you have drawn a circle around the series you were looking at when I called ''Stop.''
*Write your name across the right margin of the sheet. Put the sheet back in envelope 1, and take out the yellow sheet like this.

DIRECTIONS FOR SCRIBBLING

*Place the yellow sheet of paper before you.
I want to see how fast you can work.

*When I say "Go" make your pencil go as fast as you can back and forth down the sheet like this.
*Your speed will be measured by the number of loops you make.
*Begin well over at the left side of the page and continue until I say "Stop."
*If you come to the bottom of the page start again at the top, but do not turn over the page.
Ready—Go! (Time limit, 5 seconds.)
Stop! Turn the sheet over and try once more.
Ready—Go! (Time limit, 5 seconds.)
Stop!
Write your name across the bottom of the sheet.
*Put the sheet back in envelope 1, and take out the piece of string like this.

DIRECTIONS FOR KNOT TYING

*Straighten out your piece of string, like this.
I want to see how adept you are at tying knots.
When I say "Go!" start tying knots as quickly as you can.
*The knots must be in a row, like this.
If they are tied on top of each other they will not count.
Go as fast as you can, but try not to get the string tangled or you may lose time.
If one end gets tangled, you may begin tying from the other end.
Ready—Go! (Time limit, 1 minute.)
Stop! Do not complete tying your knot.
Put your string back in envelope 1.

The directions for scoring needed particular attention in order to make sure that the subjects would be able to follow them without confusion.

DIRECTIONS FOR SCORING

1. *Scoring of Cancellation.*

*Take the cancellation sheet out of envelope 1.
I want each one to check the number of errors on the sheet as far as the person went, that is, up to and including the series with the circle around it.
I will read across the sheet from left to right.
Check any figure which is underlined incorrectly or which is not underlined when it should be.
Be sure to check every error, including wrong underlining and omissions, up to and including the circled series.
*(Read list.)
Count the total number of checks, that is, errors made, including series wrongly underlined and omitted, and record this total on the sheet.
Count the total number of series examined. Since there are 7 series in a row, count down to the last completed row and multiply the rows by seven, and add any additional series in an incompleted row: for example, —if the person finished 4 rows and had examined 3 series on the 5th row, she would have examined 4 × 7 which is 28 series in the 4 rows, plus 3, which is 31 in all. Record this total on the sheet and subtract from it your total errors. This will give you the final score—the number of series correctly examined.
*Lift the blue sheet partly out of envelope 1 and record this final score in the score box.
*Put back the blue sheet and the cancellation sheet in envelope 1 and take out the yellow sheet.

2. *Scoring of Scribbling.*

Each trial is to be scored separately.
*Count the number of loops on the right side of the scribble.

*A real hook must be made in order to count.
Record the results of each trial on the sheet.
Lift the white sheet partly out of envelope 1 and record in the score box the score of the *best* trial.
Put back the white sheet and the yellow sheet in envelope 1 and take out the string.

3. *Scoring of Knot Tying.*

Count the number of knots completely tied.
If several knots are tied on top of each other, only count them as one knot.
Lift the pink sheet partly out of envelope 1, and record the number of knots tied in the score box.
Put back the pink sheet and the string in envelope 1.

The first directions for recording estimates were given at the beginning of the second trial. They were very specific so that the subjects would get the right idea from the start. Later on, it was possible to simplify the directions somewhat, after the subjects had had experience with estimating. The only difference in the directions for the three tasks was in the explanation of the units to be used. Where in cancellation the directions were, ''Now I want you to indicate in actual units, that is, in the number of figures correctly examined, how much you think you will gain or lose''; in scribbling we find the statement, ''This time the units are the number of loops in the best trial gained or lost over the original score''; and, in knot tying, ''This time the units are the number of knots gained or lost over the original score.''

Directions for pre-estimates and post-estimates were given, the trials coming in between. In the fourth trial the directions were varied to explain the procedure for recording the setting of the goal. The directions for cancellation will illustrate the procedure used.

Directions for Recording Estimates

Cancellation

1. *Directions for Pre-Estimates.*

*Take the cancellation sheet out of envelope 2.
Place it face down.
Again you are to underline every series containing both a 6 and a 9 as before, but first I want you to state whether or not you think you will improve on your first trial.
*For the minute, turn to the blue sheet.
*Look only at the block labeled ''Second Trial'' and disregard the rest of the sheet.
*At the top of this block you see three statements:

Shall improve.
Shall not do so well.
Score will remain the same.

All those who think they will improve on the first trial check the statement ''Shall improve.''
All those who think they won't do so well check ''Shall not do so well.''

Those who think their score will remain exactly the same check ''Score will remain the same.''

*You have your original score in the score box to refer to if necessary.

All right. Has everyone checked one of these statements?

Now, I want you to indicate in actual units, that is, in the number of groups correctly examined, how much you think you will gain or lose.

*Record these units of gain or loss in the ''Before'' column.

Those who checked ''Shall improve'' will fill in the number of additional groups of figures they *think* they will correctly examine after the words ''Think I shall gain at least,'' and the number of additional groups of figures they *may possibly* examine after the words ''May possibly gain as much as.......''

Those who checked ''Shall not do so well'' will fill in their expected loss in the same way after the words ''Think I shall lose at least.....'' and ''May possibly lose as much as......''

Those who checked ''Score will remain the same'' will fill in the two statements expressing possibility, ''May possibly gain as much as....'' and ''May possibly lose as much as......''

It is quite clear then that each person will fill in two statements?

Either ''Think I shall gain at least'' and
''May possibly gain as much as.''
Or ''Think I shall lose at least'' and
''May possibly lose as much as.''
Or ''May possibly gain as much as'' and
''May possibly lose as much as.''

All right, now turn the blue sheet over and turn to the cancellation sheet.

(Task performed)

2. *Directions for Post-Estimates.*

*You are now to indicate in the ''After'' column how much you think you actually gained or lost.

If you think you gained fill in the units of gain after the statements ''Think I gained at least.....'' and ''May possibly have gained as much as......''

If you think you lost fill in the statements ''Think I lost at least.....'' and ''May possibly have lost as much as......''

If you think your score remained exactly the same fill in the two statements expressing possibility, ''May have gained as much as.....'' and ''May have lost as much as......''

Place the blue sheet under the pink sheet.

*Take the yellow sheet out of envelope 2.

III. PERFORMANCE IN THE THREE TASKS

Before considering the subjects' estimates of improvement, we will give the results of their performance. It will be of special interest to compare improvement in the three tasks, which will be found to differ considerably in improvability.

ORIGINAL SCORES

Table 1 presents the average original scores for each of the seven groups. From this table it is seen that the averages for the different groups are not far apart.

TABLE 1

AVERAGE ORIGINAL SCORES

Group	*Av.*	*S.D.*	*Mdn.*	*Skewness*	*Range*
CANCELLATION					
Pitt	29.13	6.01	29.70	−.28	14 to 47
Sc.	26.53	5.36	26.50	+.02	14 “ 41
Arts	29.32	7.00	29.25	+.03	8 “ 51
H.S.	28.50	6.41	29.50	−.47	14 “ 41
Ind.	30.16	6.52	31.50	−.62	11 “ 45
M.M.	29.10	5.08	30.00	−.53	17 “ 39
Store	26.11	5.44	27.50	−.77	15 “ 35
SCRIBBLING					
Pitt	27.55	5.80	27.56	.00	12 to 47
Sc.	32.74	6.82	33.00	−.11	17 “ 60
Arts	27.84	6.37	28.60	−.35	12 “ 43
H.S.	29.21	5.46	29.00	+.12	20 “ 46
Ind.	34.76	9.40	34.00	+.24	20 “ 75
M.M.	28.80	5.17	29.00	−.10	21 “ 43
Store	26.28	5.39	26.00	+.21	18 “ 38
KNOT TYING					
Pitt	10.71	3.00	11.43	−.72	3 to 19
Sc.	9.89	2.75	10.30	−.45	3 “ 16
Arts	11.30	3.02	11.11	−.80	2 “ 17
H.S.	10.45	3.34	11.20	−.75	3 “ 17
Ind.	10.08	2.32	10.44	−.46	6 “ 14
M.M.	12.40	2.85	13.67	−1.34	5 “ 16
Store	11.33	2.54	11.00	+.38	3 “ 17

DISTRIBUTION OF AV. SCORES

Task	*Av.*	*S.D.*
Cancellation	28.39	1.40
Scribbling	29.60	2.80
Knot Tying	10.88	.92

The distribution of the scores shows a negative skewness in 14 out of 21 cases. That is, the scores bunched at the higher end of the scale indicating that the majority of subjects had high rather than low scores, and that the extreme cases at the high end were relatively few.

A good deal of individual variation in performance was present, as brought out by the range in scores. Taking the groups as a whole, we find in cancellation the scores ranging from 8–51; in scribbling from 12–60; and in knot tying from 2–19.

IMPROVEMENT

Improvement was measured in terms of the difference between the *original score* and performance scores in the three succeeding trials. For example, if a subject had an original score of 31, and scored 39 in the second trial, 34 in the third trial, and 40 in the fourth trial, his improvement scores would run 8, 3 and 9, respectively.

The average improvement scores for the different groups are

TABLE 2

AVERAGE IMPROVEMENT SCORES IN CANCELLATION

Group	*Av.*	*S.D.*	*Mdn.*	*Skewness*	*Range*	*% Imp.*
TRIAL II						
Pitt	9.21	5.30	8.93	+.16	−9 to 38	32
Sc.	7.63	4.05	7.33	+.22	−1 “ 21	29
Arts	9.88	4.74	10.40	−.33	−4 “ 21	34
H.S.	8.21	3.30	8.80	−.53	−1 “ 19	29
Ind.	7.16	4.80	7.66	−.31	−4 “ 21	24
M.M.	6.85	4.29	8.20	−.94	−6 “ 13	24
Store	7.22	3.91	6.50	+.55	−3 “ 19	28
TRIAL III						
Pitt	12.41	7.77	10.80	+.62	2 to 47	43
Sc.	9.47	4.35	9.00	+.32	−2 “ 24	36
Arts	10.12	4.75	11.00	−.56	−1 “ 24	35
H.S.	8.95	3.94	9.17	−.17	−1 “ 23	31
Ind.	8.55	5.35	8.75	−.11	−3 “ 24	28
M.M.	9.85	3.62	10.50	−.54	2 “ 15	34
Store	8.44	5.29	9.20	−.43	1 “ 29	32
TRIAL IV						
Pitt	14.36	7.48	14.25	+.04	1 to 58	49
Sc.	11.58	5.47	11.20	+.21	2 “ 28	44
Arts	13.27	4.96	13.78	−.31	3 “ 28	45
H.S.	12.43	4.57	12.44	−.01	4 “ 29	44
Ind.	10.89	6.04	11.00	−.05	−4 “ 27	36
M.M.	11.85	4.36	12.00	−.10	3 “ 19	41
Store	11.39	5.47	12.00	−.33	4 “ 26	44

given in Tables 2, 3 and 4. The percents are computed from the group averages.

The first thing to notice about these tables is that the averages show improvement rather than loss in all three trials. The only exception to this is the high-school group, which registers one loss of .45 or 2%. The negative skewness of the distributions indicates a fair amount of improvement by the majority. The few cases of positive skewness were brought about by a scattering of very high scores, which were not offset by equally low ones.

TABLE 3

AVERAGE IMPROVEMENT SCORES IN SCRIBBLING

Group	*Av.*	*S.D.*	*Mdn.*	*Skewness*	*Range*	*% Imp.*
			TRIAL II			
Pitt	3.04	5.54	3.53	−.27	−15 to 20	11
Sc.	3.60	7.05	5.67	−.88	−37 “ 16	11
Arts	4.55	4.96	5.36	−.49	−13 “ 19	16
H.S.	2.00	3.97	2.50	−.38	−13 “ 10	7
Ind.	4.55	7.86	6.00	−.54	−23 “ 19	13
M.M.	3.90	3.25	4.67	−.71	−5 “ 8	14
Store	2.22	4.75	4.00	−1.12	−8 “ 9	8
			TRIAL III			
Pitt	4.83	6.74	5.25	−.18	−12 to 41	18
Sc.	3.17	7.22	4.50	−.55	−33 “ 15	10
Arts	3.94	5.19	4.70	−.04	−10 “ 19	14
H.S.	−.45	4.63	1.00	−.94	−15 “ 9	−2
Ind.	3.84	6.15	4.33	−.24	−19 “ 19	11
M.M.	2.60	5.44	3.00	−.22	−8 “ 13	9
Store	1.11	4.56	1.00	+.07	−8 “ 10	4
			TRIAL IV			
Pitt	4.05	7.74	3.15	+.35	−8 to 45	15
Sc.	3.31	7.24	3.67	−.15	−26 “ 18	10
Arts	2.86	6.18	2.25	+.20	−11 “ 21	10
H.S.	.95	4.07	2.20	−.92	−8 “ 8	3
Ind.	3.53	6.52	4.50	−.45	−14 “ 20	10
M.M.	6.10	4.94	7.00	−.55	−4 “ 15	21
Store	4.61	6.53	4.67	−.03	−9 “ 23	18

In order to facilitate comparison between the three tasks, the average percents of improvement of the seven groups were averaged, and presented in graphic form in Figure 2. Cancellation shows a large gain in the second trial and definite additional gains in each of the successive trials. Scribbling, on the other hand, while

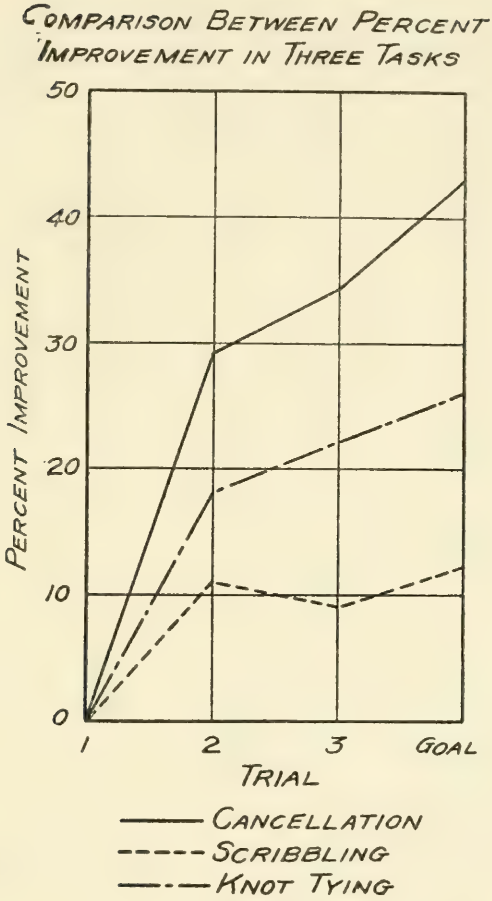
COMPARISON BETWEEN PERCENT IMPROVEMENT IN THREE TASKS

FIGURE 2

showing a moderate improvement in the second trial, does not improve noticeably beyond that point. In fact, it shows a slump in the third trial, and comes back in the fourth trial only slightly above where it was in the second trial. Knot tying shows a fairly consistent gain from trial to trial, but the percent of improvement is less than that in cancellation. In fact, the percents of improvement for all the groups are almost perfectly consistent in giving cancellation the highest percent of improvement, knot tying next, and scribbling least. (See Tables 2, 3 and 4.)

The difference in improvement between cancellation and scribbling would be expected from previous work, since cancellation is known to show rapid improvement with practice, whereas tapping,

TABLE 4

AVERAGE IMPROVEMENT SCORES IN KNOT TYING

Group	*Av.*	*S.D.*	*Mdn.*	*Skewness*	*Range*	*% Imp.*
TRIAL II						
Pitt.	1.70	2.81	2.11	−.44	−11 to 11	16
Sc.	2.00	2.99	2.18	−.18	−5 " 11	20
Arts	2.13	2.51	2.47	−.41	−5 " 10	19
H.S.	2.10	2.78	2.00	+.11	−4 " 8	20
Ind.	2.11	2.64	2.44	−.38	−7 " 8	21
M.M.	1.95	2.90	2.00	−.05	−3 " 8	16
Store	1.50	2.54	2.00	−.59	−3 " 8	13
TRIAL III						
Pitt.	2.57	3.01	3.06	−.49	−12 to 12	24
Sc.	2.03	2.73	1.94	+.10	−5 " 9	20
Arts	2.94	2.89	3.33	−.40	−3 " 10	26
H.S.	2.45	2.94	3.13	−.69	−6 " 10	23
Ind.	2.08	2.66	2.89	−.91	−5 " 8	21
M.M.	2.35	2.63	2.78	−.49	−4 " 9	19
Store	2.22	1.75	3.25	−1.77	−2 " 4	20
TRIAL IV						
Pitt.	3.03	2.93	3.50	−.48	−6 to 11	28
Sc.	2.57	2.80	2.92	−.38	−4 " 11	26
Arts	3.33	2.89	3.44	−.11	−4 " 12	29
H.S.	3.36	3.06	3.20	+.16	−2 " 9	32
Ind.	2.79	2.70	3.67	−.98	−3 " 8	28
M.M.	2.50	2.82	2.33	+.18	−1 " 7	20
Store	2.44	2.87	3.00	−.59	−4 " 9	22

which is practically the same motor process as scribbling, is likely to start in much nearer to its physiological limit.

IMPROVEMENT AS RELATED TO THE ORIGINAL SCORE

The relationship between the size of the original score and improvement is brought out in the correlation coefficients given below between absolute improvement scores and original scores.

Here we find that while size of original score and improvement correlate negatively in all three tasks, this negative correlation is much less apparent in cancellation than in either scribbling or knot tying. We can test whether this difference between cancellation and scribbling is reliable by subtracting each coefficient in scribbling from its corresponding coefficient in cancellation for all three trials, giving us 21 differences in all (*i. e.*, .02 minus —.30 = +.32; —.26 minus —.75 = +.49; —.20 minus —.56 = +.36, etc.). Averaging these differences we get an average of +.37, S.D. .15, S.D.$_{\text{av.}}$ .03, which indicates that the difference between the correlations in the two tasks is reliable. Using the same procedure, we find the difference between cancellation and knot tying to be +.40, S.D. .16, S.D.$_{\text{av.}}$ .03, again a reliable difference. The difference between scribbling and knot tying is not significant, in 11 cases the correlation coefficients for scribbling are higher than for knot tying, in 10 cases the reverse is true.

We may conclude from these results that a high original score in both scribbling and knot tying precludes the possibility of improvement to a greater extent than in cancellation. In other words, we were right in assuming that subjects with high initial scores work nearer their physiological limit in the first trial in scribbling (and knot tying) than they do in cancellation.

Correlations Between Absolute Improvement Scores and Original Scores

Group	*Pitt.*		*Sc.*		*Arts*		*H.S.*		*Ind.*		*M.M.*		*Store*	
Trial	*r*	*P.E.*	*r*	*P.E.*	*r*	*P.E.*	*r*	*P.E.*	*r*	*P.E.*	*r*	*P.E.*	*r*	*P.E.*
Cancellation														
II	.02	.06	−.26	.08	−.20	.09	−.37	.10	−.50	.09	−.24	.14	−.01	.15
III	.09	.06	−.25	.08	−.20	.09	−.35	.10	−.28	.12	−.15	.15	−.23	.14
IV	.01	.06	−.14	.08	−.20	.09	−.25	.10	−.10	.12	−.15	.15	−.10	.15
Scribbling														
II	−.30	.05	−.75	.04	−.56	.07	−.54	.08	−.67	.07	−.47	.13	−.58	.11
III	−.09	.06	−.70	.04	−.68	.06	−.61	.07	−.65	.07	−.63	.10	−.72	.08
IV	−.10	.05	−.76	.04	−.72	.05	−.58	.08	−.62	.06	−.32	.14	−.61	.10
Knot Tying														
II	−.46	.05	−.59	.06	−.65	.06	−.49	.09	−.51	.09	−.71	.08	−.69	.09
III	−.44	.05	−.56	.06	−.63	.06	−.58	.08	−.47	.10	−.58	.11	−.83	.05
IV	−.46	.05	−.59	.06	−.75	.04	−.65	.06	−.42	.10	−.60	.10	−.62	.10

IV. THE SUBJECTS' ESTIMATES

Self-estimates of improvement may first be considered as a form of behavior, without reference to its correctness.

Three kinds of estimates were provided for in this experiment: an estimate of gain; an estimate of loss; an estimate that the score was or would remain the same. When the subject recorded that his score was or would remain the same, he was in effect giving a think-estimate of zero improvement and was so counted. In the case of may-estimates, following a same response, the subject was called upon to make a statement of possible gain or loss. These may-estimates after a same response were handled separately and are discussed later in Chapter VII.

The estimates made immediately preceding the trial are referred to as pre-estimates; those following the trial as post-estimates. An estimate of gain was considered a plus quantity, an estimate of loss a minus quantity, and the average estimate for any group was based on the algebraic sum of these plus and minus quantities. The average estimates for the different groups are given in Tables 5 to 10. For purposes of comparison between tasks the last column gives also the size of these estimates in relation to the original scores, *i. e.*, it gives for each group the estimate in percent—$\frac{\text{Av. group est.}}{\text{Av. group original score.}}$ The average scores for the may-estimates represent a slightly smaller number of cases than the think-estimates, since the same-estimates have been omitted.

Estimates of Gain versus Loss

It is immediately apparent from the fact that no minus signs appear in the average estimates for any group that the subjects on the whole estimated improvement rather than loss in all three tasks.

The Distribution of the Estimates

Tendency Towards Skewness

We find the same tendency towards negative skewness in the distribution of estimates of improvement in cancellation and knot tying as we did in the original scores and actual improvement. On the whole, the subjects made a moderately high rather than a low estimate. In the few cases of positive skewness, the stragglers showed decided optimism which was not balanced by an equal

TABLE 5

Think-Estimates of Improvement in Cancellation

Group		*Av.*	*S.D.*	*Mdn.*	*Mode*	*Skewness*	*Range*	*% Est.*
				Trial II				
Pitt.........	Pre.	3.86	2.60	4.45	5	−.68	−5 to 15	13
	Post.	4.58	4.08	5.09	5	−.38	−5 " 35	16
Sc.........	Pre.	2.97	2.49	3.20	2	−.28	−4 " 15	11
	Post.	4.64	3.20	5.20	5	−.53	−3 " 20	17
Arts.........	Pre.	3.01	2.48	3.71	3	−.85	−1 " 15	10
	Post.	4.84	3.15	5.27	5	−.41	−2 " 12	17
H.S.........	Pre.	3.90	1.76	4.50	5	−1.02	0 " 10	14
	Post.	3.71	3.40	4.67	4	−.85	−10 " 10	13
Ind.........	Pre.	2.34	6.83	3.20	3	−.38	−4 " 7	8
	Post.	2.97	4.12	4.17	4	−.87	−7 " 10	10
M.M.	Pre.	2.10	1.64	3.11	3	−1.85	−3 " 8	7
	Post.	2.30	2.74	2.50	0, 5	−.22	−1 " 10	8
Store......	Pre.	3.67	2.89	3.33	3, 5	+.35	0 " 10	14
	Post.	4.00	2.45	5.00	5	−1.22	0 " 10	15
				Trial III				
Pitt.........	Pre.	4.93	3.15	5.29	3	−.34	−5 to 20	17
	Post.	5.73	4.60	5.47	5	+.17	−5 " 35	20
Sc.........	Pre.	4.18	2.76	4.67	4, 5	−.53	0 " 15	16
	Post.	5.60	3.83	5.77	5	−.13	−2 " 24	21
Arts	Pre.	3.87	2.50	4.50	5	−.76	−2 " 10	13
	Post.	4.22	3.68	4.67	5	−.37	−8 " 15	14
H.S.........	Pre.	4.83	4.08	5.09	5	−.19	0 " 27	17
	Post.	3.83	3.77	4.60	5	−.61	−8 " 10	13
Ind.........	Pre.	3.66	3.20	4.38	4	−.68	−5 " 10	12
	Post.	4.47	3.25	5.00	5	−.49	−3 " 12	15
M.M.	Pre.	2.75	3.08	3.00	2, 3	−.24	−2 " 12	9
	Post.	4.45	3.44	5.00	5	−.48	0 " 14	15
Store.....	Pre.	4.72	2.23	5.17	5	−.60	2 " 12	18
	Post.	4.56	2.91	5.25	5	−.71	−2 " 12	17
				Trial IV				
Pitt.........	Goal	10.27	6.70	9.28	10	+.44	2 to 50	35
	Post.	9.22	6.54	8.25	10	+.44	−5 " 50	28
Sc.........	Goal	9.92	4.95	10.06	10	−.08	2 " 26	37
	Post.	7.82	5.26	7.86	12	−.02	−3 " 26	29
Arts.........	Goal	8.80	4.28	8.63	10	+.12	3 " 23	30
	Post.	6.50	4.55	6.67	5	−.11	−10 " 20	22
H.S.........	Goal	9.31	4.30	9.67	10	−.21	4 " 26	33
	Post.	5.88	3.98	6.50	10	−.47	−5 " 12	21
Ind.........	Goal	8.71	3.78	8.78	5	−.06	2 " 17	29
	Post.	5.68	4.55	6.00	10	−.21	−5 " 15	19
M.M.	Goal	8.15	4.40	8.00	5	+.10	3 " 20	28
	Post.	6.70	4.36	5.50	3	+.83	0 " 18	23
Store......	Goal	8.06	3.70	9.00	10	−.76	4 " 20	33
	Post.	7.06	3.54	8.67	8	−1.36	0 " 15	27

TABLE 6

May-Estimates of Improvement in Cancellation

Group		*Av.*	*S.D.*	*Mdn.*	*Mode*	*Skewness*	*Range*	*% Est.*
				Trial II				
Pitt	Pre.	6.96	4.15	6.96	5	00	−7 to 30	24
	Post.	6.53	5.31	6.83	6	−.17	−5 “ 45	23
Sc.	Pre.	5.48	3.67	5.50	5	−.02	−7 “ 20	21
	Post.	6.63	3.91	6.79	5	−.12	−4 “ 21	25
Arts	Pre.	6.29	3.16	6.75	5	−.44	−2 “ 20	22
	Post.	7.15	3.67	7.56	6	−.34	−4 “ 17	24
H.S.	Pre.	6.59	1.96	7.17	5	−.89	0 “ 10	23
	Post.	5.72	4.38	6.73	6	−.69	−10 “ 13	20
Ind.	Pre.	6.60	2.22	6.11	5	+.20	−6 “ 12	23
	Post.	4.68	6.15	6.50	6	−.89	−10 “ 14	15
M.M.	Pre.	4.53	2.57	5.42	5	−1.04	−4 “ 8	15
	Post.	4.12	4.00	5.00	5	−.66	−3 “ 12	14
Store	Pre.	6.87	3.68	5.90	5	+.79	2 “ 13	26
	Post.	5.82	3.11	5.75	3	+.07	2 “ 15	23
				Trial III				
Pitt	Pre.	7.62	4.39	7.69	8	−.05	−5 to 30	26
	Post.	8.03	5.58	7.25	6	+.42	−3 “ 45	28
Sc.	Pre.	7.16	3.46	6.86	6	+.26	1 “ 20	27
	Post.	8.03	4.17	8.23	8	−.14	−3 “ 26	30
Arts	Pre.	6.64	3.31	7.38	8	−.67	−3 “ 16	23
	Post.	6.69	4.74	6.92	5	−.15	−10 “ 20	23
H.S.	Pre.	7.07	4.29	7.13	4	−.04	0 “ 27	25
	Post.	5.88	5.33	6.61	6	−.41	−10 “ 20	21
Ind.	Pre.	6.50	4.80	7.00	6	−.31	−8 “ 16	22
	Post.	7.11	4.50	7.50	10	−.26	−5 “ 20	24
M.M.	Pre.	4.10	4.37	4.25	3	−.10	−4 “ 14	14
	Post.	7.00	4.18	6.50	4	+.36	0 “ 17	24
Store	Pre.	6.67	2.58	6.60	6	+.08	4 “ 15	26
	Post.	6.39	3.65	6.50	6	−.09	−3 “ 15	24
				Trial IV				
Pitt	Goal Post.	10.19	7.44	10.17	10	+.01	−6 to 50	35
Sc.	Goal Post.	10.28	6.02	10.25	7, 12, 15	+.01	−7 “ 28	39
Arts	Goal Post.	8.67	5.64	8.91	10	−.13	−15 “ 25	30
H.S.	Goal Post.	7.79	4.76	8.78	10	−.62	−5 “ 15	27
Ind.	Goal Post.	8.35	6.06	9.50	3, 12	−.57	−10 “ 20	28
M.M.	Goal Post.	9.00	4.80	8.00	5	+.63	0 “ 20	30
Store	Goal Post.	9.00	4.16	10.10	10	−.79	2 “ 20	35

TABLE 7

Think-Estimates of Improvement in Scribbling

Group		*Av.*	*S.D.*	*Mdn.*	*Mode*	*Skewness*	*Range*	*% Est.*
				Trial II				
Pitt	Pre.	6.02	3.67	5.61	5	+.34	−3 to 20	22
	Post.	5.25	4.61	5.60	5	−.23	−10 " 20	19
Sc	Pre.	5.13	3.06	5.15	5	−.02	0 " 15	16
	Post.	4.03	5.75	4.78	5	−.39	−20 " 20	12
Arts	Pre.	4.75	2.76	5.22	5	−.51	−2 " 15	17
	Post.	4.08	4.02	4.38	0, 3, 10	−.22	−6 " 15	15
H.S.	Pre.	5.45	2.46	5.60	5	−.18	0 " 10	19
	Post.	5.67	4.31	5.60	5	+.05	−6 " 15	19
Ind.	Pre.	5.79	3.37	5.60	5	+.17	0 " 15	17
	Post.	3.74	4.89	5.20	5	−.90	−5 " 15	11
M.M.	Pre.	3.25	2.05	3.67	2	−.61	0 " 8	11
	Post.	4.40	3.50	4.00	3	+.34	−2 " 10	15
Store	Pre.	5.39	3.25	5.40	5	−.01	0 " 10	21
	Post.	4.56	4.44	5.00	5	−.30	−5 " 15	17
				Trial III				
Pitt	Pre.	5.72	4.66	5.57	5	+.10	−5 to 24	21
	Post.	5.16	5.34	5.14	5	+.01	−10 " 30	15
Sc	Pre.	5.47	4.29	5.31	5	+.11	0 " 20	17
	Post.	4.78	5.59	5.06	5	−.15	−10 " 30	15
Arts	Pre.	5.05	4.25	5.00	6	+.04	−1 " 20	18
	Post.	4.39	4.01	4.00	5	+.27	0 " 20	16
H.S.	Pre.	5.86	3.39	5.83	3	+.03	0 " 15	20
	Post.	5.38	3.29	5.50	5	−.11	0 " 13	18
Ind.	Pre.	5.16	4.81	5.56	5	−.25	−10 " 20	15
	Post.	2.82	4.84	5.00	5	−1.35	−10 " 10	8
M.M.	Pre.	5.00	3.74	5.00	5	00	−3 " 15	17
	Post.	5.10	4.31	4.50	2	+.04	0 " 15	18
Store	Pre.	6.94	5.22	5.50	3	+.83	0 " 20	26
	Post.	5.11	4.08	5.00	5	+.08	0 " 18	19
				Trial IV				
Pitt	Goal	10.11	5.93	10.30	10	−.10	0 to 40	37
	Post.	6.80	5.63	7.40	10	−.03	12 " 25	25
Sc	Goal	10.08	7.08	10.16	10	−.03	0 " 50	31
	Post.	6.81	6.21	6.20	5	+.30	−5 " 30	21
Arts	Goal	8.61	4.90	8.46	10	+.09	0 " 30	31
	Post.	4.33	5.99	5.00	5	−.34	−30 " 16	16
H.S.	Goal	9.26	4.28	9.80	10	−.38	3 " 25	32
	Post.	7.05	3.93	6.00	5, 10	+.80	1 " 20	24
Ind.	Goal	9.68	5.08	9.36	10	+.19	1 " 20	28
	Post.	5.95	5.13	5.64	5	+.18	−10 " 15	17
M.M.	Goal	8.95	5.40	7.00	6	+1.08	4 " 27	31
	Post.	6.00	5.41	5.33	4	+.37	−5 " 20	21
Store	Goal	9.89	4.34	10.29	10	−.28	3 " 20	38
	Post.	9.00	6.19	7.50	5	+.73	2 " 30	34

TABLE 8

May-Estimates of Improvement in Scribbling

Group		*Av.*	*S.D.*	*Mdn.*	*Mode*	*Skewness*	*Range*	*% Est.*
Trial II								
Pitt	Pre.	9.57	5.28	8.75	10	−.10	−8 to 30	35
	Post.	8.17	6.34	7.80	10	−.30	−15 “ 30	30
Sc.	Pre.	9.37	3.71	9.10	10	−.59	3 “ 20	29
	Post.	5.82	7.65	7.10	10	−.89	−25 “ 30	22
Arts	Pre.	8.21	3.97	7.25	10	−.03	−3 “ 20	30
	Post.	6.57	6.22	6.25	10	−.33	−15 “ 20	23
H.S.	Pre.	8.61	3.19	8.75	10	−1.07	2 “ 15	30
	Post.	9.02	5.82	9.22	10	−.62	−10 “ 20	31
Ind.	Pre.	10.26	4.25	8.32	10	−.04	3 “ 20	26
	Post.	6.86	8.50	6.75	10	−.28	−15 “ 20	19
M.M.	Pre.	6.47	2.77	5.25	3, 5, 8, 10	+.13	3 “ 12	23
	Post.	7.33	4.98	6.00	5	−1.61	−3 “ 18	25
Store	Pre.	9.06	4.51	8.00	5, 10	−.63	3 “ 20	34
	Post.	7.63	6.32	5.67	6, 10	+.46	−5 “ 25	29
Trial III								
Pitt	Pre.	8.89	6.33	7.82	10	+.03	−15 to 30	32
	Post.	7.96	5.65	7.14	10	−.10	−15 “ 40	29
Sc.	Pre.	9.69	5.59	9.25	15	+.37	3 “ 30	30
	Post.	8.65	8.21	9.04	10	−.51	−15 “ 40	27
Arts	Pre.	8.76	5.50	7.50	8	+.09	−3 “ 30	32
	Post.	7.95	4.94	6.00	10	+.58	2 “ 25	28
H.S.	Pre.	9.20	3.99	8.11	10	−.68	2 “ 20	31
	Post.	8.12	3.91	7.17	10	−.04	2 “ 15	28
Ind.	Pre.	9.53	7.09	8.41	10	−.37	−15 “ 25	29
	Post.	5.86	8.37	7.50	10	−.95	−13 “ 20	17
M.M.	Pre.	7.70	5.00	6.50	5	+.12	−4 “ 20	27
	Post.	7.84	5.78	5.75	15	+.57	1 “ 25	28
Store	Pre.	11.35	6.93	8.25	15	+.48	2 “ 25	44
	Post.	8.94	6.17	7.20	8	+.34	3 “ 30	34
Trial IV								
Pitt	Goal Post.	9.22	7.33	9.10	10	−.37	−15 to 30	33
Sc.	Goal Post.	10.62	7.32	9.38	10	+.10	−10 “ 35	30
Arts	Goal Post.	7.73	7.58	7.22	8	−.19	−32 “ 20	28
H.S.	Goal Post.	9.76	4.68	8.00	7, 12	−.15	2 “ 22	33
Ind.	Goal Post.	9.56	7.31	8.44	10	−.36	−15 “ 25	27
M.M.	Goal Post.	9.74	6.95	8.33	10	−.25	−5 “ 25	34
Store	Goal Post.	12.28	7.25	8.60	10	+.70	3 “ 35	47

TABLE 9

THINK-ESTIMATES OF IMPROVEMENT IN KNOT TYING

Group		*Av.*	*S.D.*	*Mdn.*	*Mode*	*Skewness*	*Range*	*% Est.*
TRIAL II								
Pitt	Pre.	2.54	1.83	2.96	2	−.69	−7 to 11	24
	Post.	1.49	2.47	2.14	2	−.79	−9 “ 8	14
Sc.	Pre.	1.69	1.18	2.19	2	−1.27	4 “ 7	17
	Post.	1.18	2.36	1.76	1	−.74	−6 “ 7	12
Arts	Pre.	1.44	1.59	2.05	2	−1.15	−2 “ 8	11
	Post.	.50	2.40	1.27	1	−.92	−5 “ 8	4
H.S.	Pre.	2.75	2.50	3.30	2, 3	−.65	−5 “ 10	26
	Post.	1.38	3.59	2.11	1, 2	−.61	−10 “ 10	13
Ind.	Pre.	1.61	1.40	2.29	2	−1.46	−2 “ 5	16
	Post.	1.66	2.14	2.32	2	−.93	−6 “ 5	16
M.M.	Pre.	1.65	1.49	1.80	1	−.30	0 “ 5	13
	Post.	1.10	2.62	2.00	2	−1.03	−4 “ 5	9
Store	Pre.	1.69	1.85	1.60	0, 2	+.14	−2 “ 5	17
	Post.	1.67	2.24	2.50	2, 3	−1.11	−4 “ 5	15
TRIAL III								
Pitt	Pre.	2.30	2.31	2.73	2	−.56	−7 to 10	21
	Post.	1.55	2.71	2.24	2	−.76	−10 “ 8	14
Sc.	Pre.	1.99	2.14	2.48	2	−.69	−7 “ 10	20
	Post.	1.18	2.63	1.58	0	−.46	−8 “ 6	12
Arts	Pre.	1.19	2.05	1.71	1	−.76	−5 “ 8	11
	Post.	.89	2.64	1.42	0	−.60	−6 “ 8	8
H.S.	Pre.	2.48	2.20	2.67	2	−.26	−2 “ 8	24
	Post.	1.21	2.91	2.25	2	−1.07	−6 “ 6	12
Ind.	Pre.	2.03	1.60	2.46	2	−.81	−2 “ 10	21
	Post.	1.45	2.50	2.50	2, 3	−1.26	−7 “ 5	14
M.M.	Pre.	1.80	1.60	2.25	2	−.84	0 “ 5	15
	Post.	.60	2.15	1.00	0	−.56	−5 “ 5	5
Store	Pre.	2.83	2.22	4.00	5	−1.58	−1 “ 6	25
	Post.	1.39	2.54	1.78	1	−.46	−6 “ 6	12
TRIAL IV								
Pitt	Goal	4.46	2.65	4.78	5	−.36	−9 to 16	42
	Post.	2.53	2.78	3.07	2	−.58	−8 “ 10	24
Sc.	Goal	4.15	2.44	4.30	5	−.18	1 “ 14	42
	Post.	2.23	2.60	2.62	2, 3	−.41	−3 “ 10	23
Arts	Goal	3.53	2.03	3.52	3	+.01	0 “ 10	31
	Post.	1.77	2.56	2.15	1, 2	−.45	−5 “ 12	16
H.S.	Goal	5.10	3.37	4.75	5	+.31	1 “ 16	49
	Post.	2.88	2.82	3.25	3	−.39	−2 “ 10	28
Ind.	Goal	4.39	2.53	4.44	5	−.05	1 “ 12	44
	Post.	2.26	2.40	3.00	5	−.93	−4 “ 5	22
M.M.	Goal	3.55	1.66	3.67	3	−.22	1 “ 8	29
	Post.	1.30	2.19	1.50	0	−.27	−2 “ 6	10
Store	Goal	3.89	1.94	5.00	5	−1.72	0 “ 8	34
	Post.	2.78	2.35	3.33	2	−.70	−5 “ 8	25

TABLE 10

May-Estimates of Improvement in Knot Tying

Group		*Av.*	*S.D.*	*Mdn.*	*Mode*	*Skewness*	*Range*	*% Est.*
				Trial II				
Pitt.........	Pre.	4.45	2.54	5.04	4, 5	−.70	−12 to 12	42
	Post.	2.32	3.66	4.52	4	−.18	−13 “ 10	22
Sc...........	Pre.	3.73	2.37	4.16	3, 4	−.54	−5 “ 10	39
	Post.	2.32	3.62	3.58	3	−1.04	−9 “ 10	23
Arts.........	Pre.	3.51	2.54	4.14	3	−.50	−4 “ 10	35
	Post.	1.09	3.78	2.67	2	−1.25	−10 “ 10	10
H.S..........	Pre.	5.31	3.57	5.56	4, 5	−.21	−5 “ 15	52
	Post.	2.18	5.36	3.58	2	−.78	−11 “ 14	21
Ind..........	Pre.	3.84	2.52	4.79	5	−1.13	−4 “ 10	37
	Post.	3.03	3.19	3.92	3	−.84	−7 “ 8	30
M.M.	Pre.	4.07	1.62	4.33	2, 4, 6	−.48	2 “ 7	33
	Post.	2.00	3.68	4.00	4	−1.63	−6 “ 7	16
Store......	Pre.	4.85	3.37	5.25	4	−.36	−4 “ 10	43
	Post.	2.87	3.26	3.78	3	−.84	−5 “ 7	25
				Trial III				
Pitt.........	Pre.	4.06	3.52	4.90	3	−.72	−12 to 12	38
	Post.	2.57	4.04	3.83	3	−.94	−14 “ 10	24
Sc...........	Pre.	3.73	2.37	4.42	3, 4	−.44	−10 “ 12	39
	Post.	1.98	4.29	3.50	3	−1.06	−12 “ 9	20
Arts.........	Pre.	2.51	3.63	3.75	4	−1.02	−10 “ 10	22
	Post.	1.67	4.07	3.18	3, 4	−1.11	−7 “ 10	14
H.S..........	Pre.	4.26	3.62	4.86	3	−.50	−9 “ 12	41
	Post.	2.38	4.86	3.50	2	−.69	−8 “ 15	22
Ind..........	Pre.	4.33	2.64	4.43	3	−.11	−4 “ 10	43
	Post.	2.86	4.06	4.25	3, 5	−1.03	−9 “ 10	29
M.M.	Pre.	4.38	1.39	4.83	3, 4, 5	−.97	2 “ 7	37
	Post.	1.57	3.42	3.50	4	−.79	−6 “ 5	13
Store......	Pre.	5.20	3.49	6.75	2	−1.33	−3 “ 10	45
	Post.	2.47	3.66	2.92	2	−.37	−7 “ 7	22
				Trial IV				
Pitt.........	Goal							
	Post.	3.84	3.55	4.94	5	−.93	−10 to 11	38
Sc...........	Goal							
	Post.	3.58	3.61	4.35	5	−.64	−5 “ 12	36
Arts.........	Goal							
	Post.	2.98	3.47	3.95	4	−.84	−5 “ 14	27
H.S..........	Goal							
	Post.	4.55	3.88	5.00	4	−.35	−4 “ 14	43
Ind..........	Goal							
	Post.	3.78	3.62	5.00	5, 6	−1.01	−6 “ 10	38
M.M.	Goal							
	Post.	2.67	3.56	3.75	5	−.91	−3 “ 8	22
Store......	Goal							
	Post.	4.17	3.08	5.25	2	−1.05	−5 “ 8	37

tendency towards pessimism. In scribbling, the think-estimates show about an equal division of positive and negative skewness, although the may-estimates show a preponderance of negative cases (23 out of 34).

Tendency to Estimate in Round Numbers

A study of the modes brings out a tendency to estimate in round numbers in scribbling and cancellation, for the estimates of 5 and 10 occur in these tasks with a frequency that can scarcely be due to chance. This same kind of set was found in the extreme cases in which the subjects estimated in tens, as 20, 30 and 40.

Difference in Range

The range of estimates for the different groups brings out the presence of a good deal of individual variation. If we compare the range of the think-estimates and the may-estimates, however, we do not find so wide a divergence as might be expected. Taking the groups as a whole, we find the may-estimates ranging 5 points below and 0 points above the think-estimates in cancellation, 2 points below and 10 points above the think-estimates in scribbling, and 4 points below and 2 points above the think-estimates in knot tying.

CHARACTERISTICS OF THE DIFFERENT ESTIMATES

The main characteristics of the different estimates are shown in Figure 3, in which the average percent estimates for the seven groups are given. They may be summarized as follows:

1. The may-estimates exceed the think-estimates in all three tasks, although the two types of estimates were not found to differ greatly in regard to range.
2. The post-estimates are lower than the pre-estimates in scribbling and knot tying, but not in cancellation.
3. There is little difference between the estimates in the second and third trials. Setting the goal, however, resulted in an abrupt rise in all three tasks. This rise is also reflected in the post-estimates in the fourth trial, particularly in knot tying.
4. The pre-estimate percents are decidedly lower in cancellation than in the other two tasks, although this difference mostly disappears in setting the goal. These low pre-estimates of improvement in cancellation go along with large actual improvement, and thus give large under-estimation, as will be found later.
5. The post-estimates in knot tying show a marked drop as compared with the pre-estimates, except in the fourth trial, and fail to

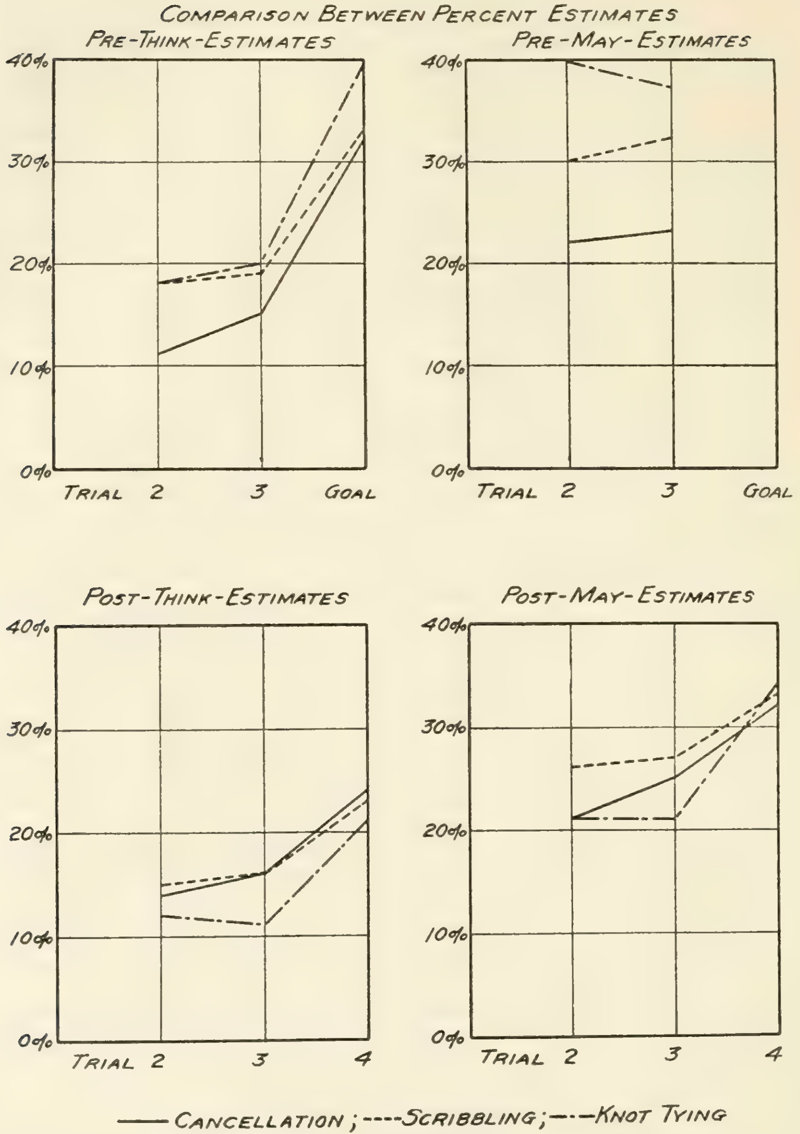

FIGURE 3

show the rise from the second to the third trial characteristic of the post-estimates in the other tasks.

From these results, the subjects appear to have been fairly optimistic about improvement in all three tasks. To judge from the difference between pre- and post-estimates, they seem to have compensated in post-estimates for possible under-estimation in cancellation in pre-estimates, and over-estimation in scribbling and knot tying. Whether this compensation was justified will be apparent when the estimates are studied in relation to actual improvement.

ESTIMATES AS RELATED TO THE ORIGINAL SCORE

In discussing this subject of estimates considered apart from actual improvement, the rôle of the original score must not be lost sight of, for it was probably largely instrumental in setting the size of the estimates. We can get an idea of the influence of the size of the original score by comparing the percent estimates of those standing in the first quartile in original score (Q1) with the percent estimates of those in the lowest quartile (Q4).

Using the Pitt. group as a representative group, it being the largest, and pre-think-estimates and goal as typical of the estimates as a whole, we find from Table 11 that these subjects did take into consideration the size of the original score in estimating. In each trial, in all three tasks, the percent estimates of those having the highest original scores were lower than the percent estimates of those with the lowest scores. The percent improvement was also lower for those with the highest original scores.

TABLE 11

PERCENT PRE-THINK ESTIMATES AND PERCENT IMPROVEMENT OF THE HIGHEST AND LOWEST QUARTILE IN ORIGINAL SCORE

(PITT. GROUP)

	Cancellation		*Scribbling*		*Knot Tying*	
	% Est.	*% Imp.*	*% Est.*	*% Imp.*	*% Est.*	*% Imp.*
TRIAL II						
Q1	09	24	19	09	14	05
Q4	20	38	31	26	41	51
TRIAL III						
Q1	13	37	18	09	15	10
Q4	25	53	24	32	48	64
GOAL						
Q1	28	40	31	11	27	09
Q4	56	76	49	32	76	64

V. ESTIMATES AS RELATED TO IMPROVEMENT

How do the subjects' estimates of improvement compare with actual improvement?

In examining Figure 4, in which the comparison between percent estimates and percent improvement is presented graphically, striking differences in the three tasks appear. Estimated improvement falls decidedly below actual improvement in cancellation, decidedly above in scribbling, and neither above nor below on the whole in knot tying.

In cancellation we get a steady rise both in improvement and in estimates from the second to the fourth trial. In scribbling there is no additional improvement to amount to anything after the second trial, but the estimates show a steady rise. In knot tying there is a slight improvement from trial to trial but the estimates do not follow this rise consistently. In all three tasks the fourth trial, in which the goal was set, brings about a decided increase in estimates. This increase in estimates, however, is decidedly greater than the increase in improvement. In other words, the setting of the goal did not bring about marked improvement in performance.

Over- and Under-Estimation

While the general tendency towards over- and under-estimation is brought out in Figure 4, the actual amount of constant error in estimation is computed by subtracting actual improvement from estimated improvement with regard to sign. Thus, if one subject improved 7 and and estimated a loss of 2; and a second subject improved 10 and estimated a gain of 15, we would have an under-estimation of 9 in one case and an over-estimation of 5 in the other, which would give us an average constant error of —2 (—4 divided by 2). The average constant errors for each group were computed and the percent of error found by dividing the average constant errors for each group by the average original scores for each group. The average percent of over- and under-estimation for the groups combined is presented in Table 12.

Significance of Difference Found in Constant Error

Figure 4 brought out the fact that under-estimation occurred in the task showing the highest percentage of improvement, cancellation, and over-estimation in the task in which least improvement was made, scribbling. That this difference was not due to chance

COMPARISON BETWEEN PERCENT ESTIMATES AND PERCENT IMPROVEMENT IN THREE TASKS

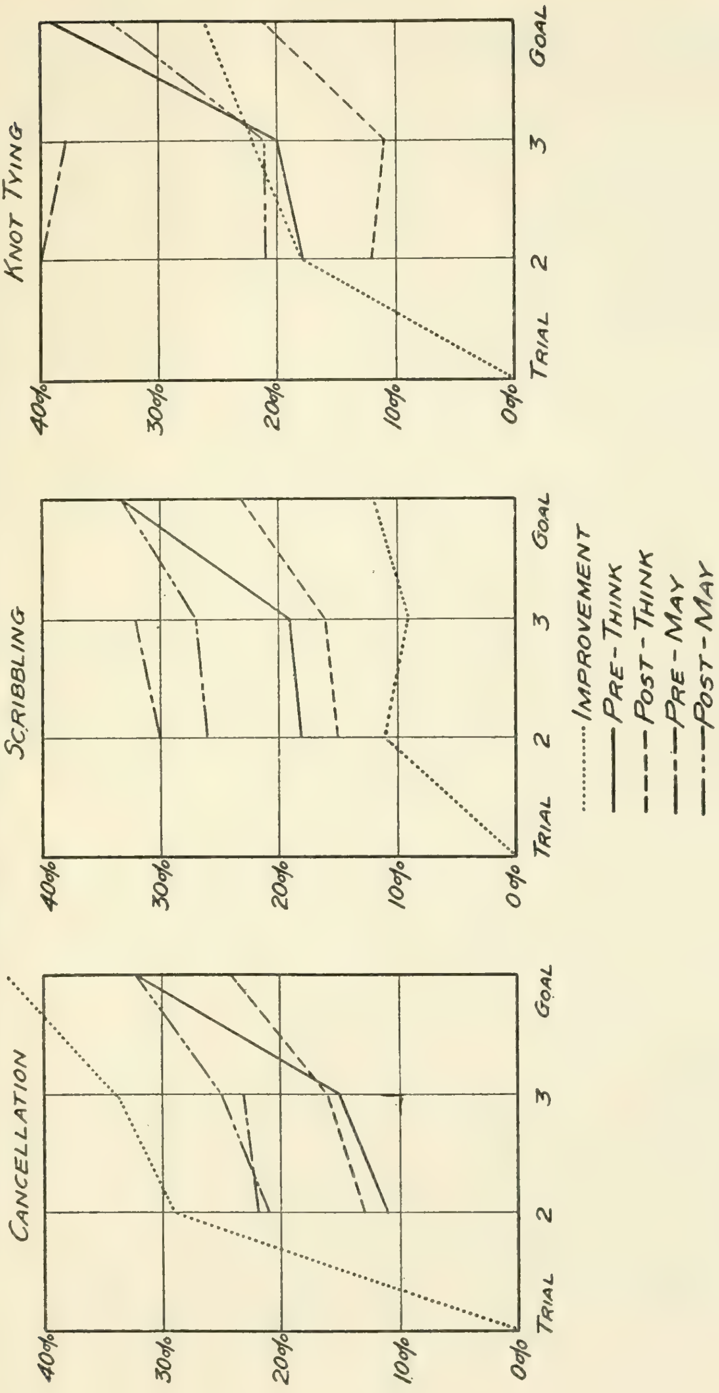

FIGURE 4

TABLE 12

Average Percent Over- and Under-Estimation for Combined Groups

Trial	*Cancellation* %	*Cancellation* *S.D.*	*Scribbling* %	*Scribbling* *S.D.*	*Knot Tying* %	*Knot Tying* *S.D.*
Think Before II	−17.1	3.05	6.1	5.22	.1	5.11
Think Before III	−19.2	4.43	10.0	7.76	−2.6	5.68
Think Before II & III	−18.2	4.15	8.1	2.59	−1.2	5.57
May Before II	−6.9	3.80	16.6	4.91	21.3	7.71
May Before III	−11.4	4.69	22.6	8.55	15.6	8.81
May Before II & III	−9.1	4.84	19.6	7.59	18.4	7.79
Think After II	−14.6	2.12	4.1	5.44	−5.6	4.68
Think After III	−17.9	3.09	6.9	7.90	−10.9	3.87
Think After II & III	−16.2	3.15	5.5	6.84	−8.3	4.87
May After II	−7.4	2.28	13.3	5.95	2.5	6.99
May After III	−10.2	3.18	15.7	6.54	−2.4	5.70
May After II & III	−8.8	3.65	14.5	6.37	0.0	6.85
Goal	−12.3	3.22	19.9	4.97	12.1	5.00
Think After IV	−19.0	3.16	10.1	6.60	−5.7	4.94
May After IV	−11.0	3.77	20.7	6.02	6.1	4.41

becomes clear in scrutinizing the percentages in Table 12. Consider the signs for cancellation—they are minus in every instance. Even the may-estimates fall below actual improvement by a substantial amount. In scribbling, we find a strong preponderance of over-estimation as shown by the positive signs throughout. In knot tying, on the other hand, there is about an even division between over- and under-estimation, the general average for all estimates coming out at about 3% for over-estimation.

Constant Error of First Estimates Compared with That of Pre-Estimates in Following Trials

How do the estimates of improvement made before the subjects have had an opportunity to note their actual improvement compare with their pre-estimates in later trials?

This question was answered by comparing the percent of over- and under-estimation (*i. e.*, constant error) in the pre-estimates of the second trial with that in the third and fourth trials. In computing the differences the percent value for each group of subjects was taken as a unit, and the difference between the error in two estimates was found from these values. The average, S.D. and S.D.$_{\text{av.}}$ for this set of differences were then found. For example, to answer the question whether the pre-estimates in Trial II in cancellation showed greater or less under-estimation than the pre-estimates in Trial III, we took from the original data the seven differences covered by the formula "Think Before II minus Think Before III." The average of these differences was —2.1, S.D. 4.05, S.D.$_{av.}$ 1.35, which indicated that the subjects' first estimates showed slightly less under-estimation than pre-estimates in the third trial. Since these differences represent averages of small samples (7 items in the case of one trial, 14 in the case of differences between estimates in two trials), the significance of the means was tested by the use of the T-formula described by Fisher in "Statistical Methods for Research Workers," pages 106–112. In the tables that follow T will be listed directly under the S.D.$_{av.}$ so that it will be possible at a glance to estimate the reliability of the means when the smallness of the sample is taken into consideration. For a difference to be considered reliable T must be at least 3.

TABLE 13
AVERAGE PERCENT DIFFERENCE BETWEEN OVER- AND UNDER-ESTIMATION IN PRE-ESTIMATES TRIAL II AND PRE-ESTIMATES TRIAL III AND GOAL

Trial	*Cancellation*			*Scribbling*			*Knot Tying*		
	Av.% Diff.	*S.D.*	*S.D.$_{av.}$*	*Av.% Diff.*	*S.D.*	*S.D.$_{av.}$*	*Av.% Diff.*	*S.D.*	*S.D.$_{av.}$*
Think Before II–	−2.1	4.05	1.53	−3.9	6.24	2.35	*2.7	5.26	1.98
Think Before III....			T = 1.27			T = 1.53			T = 1.26
May Before II–	−4.6	4.37	1.65	−6.0	9.23	3.48	5.7	5.80	2.18
May Before III......			T = 2.58			T = 1.59			T = 2.40
Think Before II–	4.9	3.68	1.39	−13.7	3.41	1.29	−12.0	5.40	2.04
Goal..........................			T = 3.27			T = 9.82			T = 5.45
May Before II–	−5.4	4.75	1.79	−3.3	3.01	1.14	9.1	6.20	2.34
Goal..........................			T = 2.78			T = 2.68			T = 3.60

*Percent of over-estimation in knot tying in think-before-II-estimates: .1%, under-estimation in think-before-III-estimates: −2.6%; difference between II and III: 2.7%.

From Table 13 we conclude that in cancellation and scribbling the subjects under- or over-estimated possible improvement before trial to a less extent when they had performed the task but once, than when they had repeated it a second or third time. This exaggeration of the constant error in later trials was brought about by the subjects' tendency to increase their estimates of improvement moderately from trial to trial, regardless of performance. Thus the marked improvement from the second to the fourth trial in cancellation and the lack of improvement in scribbling after the second trial (see Figure 4) increased the discrepancy between estimates and improvement as the experiment progressed. [The exception to this was the goal trial in cancellation, which lessened the difference by raising the (otherwise too low) estimate.]

We find in knot tying that the subjects tended to reduce their estimates after the first trial. Even the goal was not so large as the may-estimate in the first trial. This suggests that the subjects felt that they had over-estimated the first time. In other words, they depended more on actual performance in making their later estimates in knot tying than in the other two tasks. Since a moderate increase in improvement was made in knot tying, this tendency to correct for possible over-estimation resulted in under-estimation in some instances.

Constant Error in Post-Estimates Compared with That in Pre-Estimates

When we compare the constant errors of pre- and post-estimates in knot tying, we find the tendency to correct for over-estimation cropping up again in the post-estimates, for in all three trials the post-estimates show either greater under-estimation or less over-estimation than do the pre-estimates. (See Table 14.)

TABLE 14

Average Percent Difference Between Over- and Under-Estimation in Pre- and Post-Estimates

Trial	*Cancellation*			*Scribbling*			*Knot Tying*		
	Av.% Diff.	*S.D.*	*S.D.*$_{av.}$	*Av.% Diff.*	*S.D.*	*S.D.*$_{av.}$	*Av.% Diff.*	*S.D.*	*S.D.*$_{av.}$
Think Before–Think After (II & III)	2.0	2.85	.76	2.6	3.09	.83	−7.1	4.10	1.10
			T = 2.53			T = 3.08			T = 6.40
May Before–May After (II & III)	.4	3.85	1.03	5.1	6.25	1.67	18.4	4.88	1.30
			T = .37			T = 2.95			T = 13.63
Goal–Think After IV	−6.7	3.33	1.26	9.7	3.53	1.37	17.9	3.72	1.40
			T = 4.92			T = 6.74			T = 11.76
Goal–May After IV	1.3	4.20	1.58	−.9	4.05	1.53	6.0	3.67	1.38
			T = .76			T = .55			T = 4.00

Table 14 also shows a tendency to correct in the post-estimates the constant error in the pre-estimates, in cancellation and scribbling. The only exception to this is in the case of the goal estimate, which in cancellation shows decidedly less under-estimation than the post-think-estimate, and in scribbling slightly less over-estimation than the post-may-estimate.

Distribution of Constant Errors

So far we have been considering only the group tendency towards over- or under-estimation. When the distribution of indi-

DISTRIBUTION OF CONSTANT ERRORS OF ESTIMATION
PRE-THINK-ESTIMATES, TRIAL 3
(406 CASES)

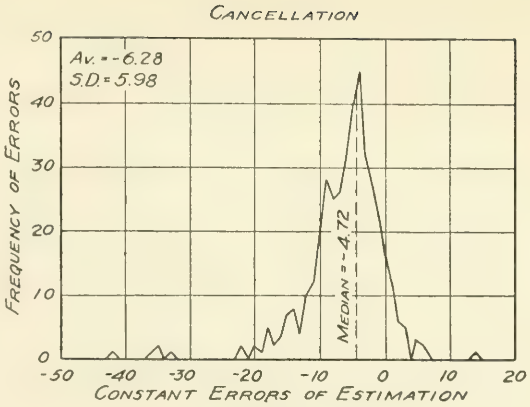

FIGURE 5

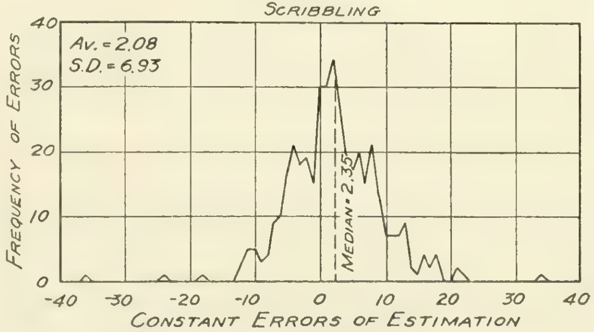

FIGURE 6

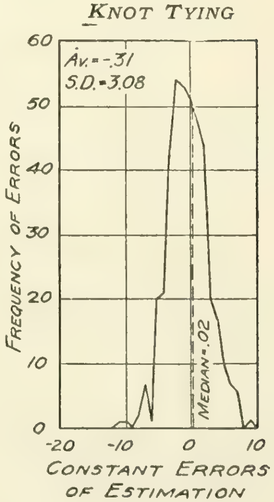

FIGURE 7

vidual errors is plotted, as in Figures 5, 6 and 7, which show a form of distribution typical for each of the three tasks, it is seen at once that the distribution is unimodal, with no gap between the positive and negative errors. The distributions appear fairly normal to the eye, but computation of skewness according to the formula,

Skewness 3 (mean—median) / S.D.,

shows in many cases a significant degree of asymmetry, as follows:

		Cancellation	*Scribbling*	*Knot Tying*
Trial II	Pre.	−40	+.04	−53
	Post.	−58	−29	−66
Trial III	Pre.	−78	−12	−32
	Post.	−73	−23	−72
Trial IV	Goal	−43	−31	−14
	Post.	−79	−18	−36

The skewness is with one exception negative, the errors being massed at the high end. In cancellation the extreme errors were more likely to lie on the negative side, whereas in scribbling they frequently appeared on both sides.

Extreme Errors of Over- and Under-Estimation

The extreme instances of over- or under-estimation are sometimes rather curious. One individual with the high initial score in scribbling of 60 pre-estimated for the second trial a modest gain of 2, but actually lost 37; his post-estimate, though allowing for a loss of 20, was still far above his actual performance. Another subject with an initial score of 75 lost 23 in Trial 3 but post-estimated a small gain of 5, giving him an over-estimate of 28. Large errors of over-estimation, then, were not always due to great optimism, but sometimes to modest estimates along with actual losses in performance. The case of greatest under-estimation in knot tying followed an original score of 14 and a gain of 5 which was post-estimated to be a loss of 10—an estimate which must have been an expression of a feeling of failure, rather than a real estimate, since it would imply that the subject thought she had made only 4 knots, whereas she had actually made 19! In cancellation extreme errors seem to have been due in many cases to bad guesses, since they had little relation to the size of the original score.

On the whole, however, individuals making high original scores in any task tended to over-estimate their improvement more than individuals with low original scores. This fact is brought out in

TABLE 15

PERCENT OF OVER- AND UNDER-ESTIMATION IN PRE-THINK-ESTIMATES OF THE HIGHEST AND LOWEST QUARTILES IN ORIGINAL SCORE

	Pitt. %	*Sc.* %	*Arts* %	*Ind.* %	*H.S.* %	*M.M.* %	*Store* %
CANCELLATION							
TRIAL II							
Q1	−15	−16	−15	−09	−15	−10	−12
Q4	−18	−24	−33	−27	−28	−21	−17
Q1 − Q4	03	08	18	16	13	11	05
TRIAL III							
Q1	−24	−19	−14	−18	−13	−21	−08
Q4	−28	−34	−27	−25	−28	−32	−26
Q1 − Q4	04	15	13	07	15	11	18
GOAL							
Q1	−12	−07	−13	−09	−10	−16	−11
Q4	−20	−08	−21	−10	−30	−19	−11
Q1 − Q4	08	01	08	01	20	03	00
SCRIBBLING							
TRIAL II							
Q1	10	16	08	09	17	00	12
Q4	05	−13	−15	−11	03	−10	03
Q1 − Q4	05	29	23	20	14	10	09
TRIAL III							
Q1	09	18	13	05	28	11	23
Q4	−08	−02	−15	00	12	−12	11
Q1 − Q4	17	20	28	05	16	23	12
GOAL							
Q1	20	26	29	18	33	10	32
Q4	17	02	08	12	20	−06	−05
Q1 − Q4	03	24	21	06	13	16	37
KNOT TYING							
TRIAL II							
Q1	09	06	03	06	16	16	15
Q4	−10	−26	−37	32	−16	−33	−19
Q1 − Q4	19	32	40	38	32	49	34
TRIAL III							
Q1	05	05	01	14	12	13	10
Q4	−16	−05	−63	−26	−18	−26	02
Q1 − Q4	21	10	64	40	30	39	08
GOAL							
Q1	18	22	11	18	28	25	08
Q4	12	09	−34	00	16	−48	23
Q1 − Q4	06	13	45	18	12	73	−15

Cancellation*: Av. diff. 9.4, S.D. 5.78, S.D.$_{av.}$ 1.04.
Scribbling*: Av. diff. 16.7, S.D. 8.75, S.D.$_{av.}$ 1.91.
Knot Tying*: Av. diff. 29.0, S.D. 6.32, S.D.$_{av.}$ 1.38.

---

*Average for combined groups.

Table 15, which compares the highest quartile (Q1) and the lowest quartile (Q4), divided according to original scores. In 61 instances out of the 63 in the table, the high quartile either over-estimated more or under-estimated less than the low quartile. The average of the 21 differences for each task, as given at the bottom of the table, is significantly positive in each instance.

Thus it is certain that individuals with high initial scores tend to over-estimate their subsequent improvement, while individuals with low initial scores tend to under-estimate improvement. This is true in spite of the fact shown in Table 11, namely, that individuals with high initial scores tend to estimate relatively smaller gains than those with low initial scores. The meaning is that the high-scoring individuals, while estimating relatively small gains, actually gain still less; and that the low-scoring individuals, while estimating relatively large gains, actually gain still more.

Accuracy of Estimates

The question of the accuracy of estimation can be separated from that of the tendency to over- or under-estimate. Under the head of "accuracy" the question is as to how closely a subject's estimates of his improvement approximated to the actual improvement in the trial concerned. The answer to this question is found in the *mean error of estimate.*

The mean error was computed by averaging the differences between improvement and estimates of improvement, this time without regard to sign—that is, disregarding the fact of whether each estimate of improvement was greater or less than actual improvement. For example, if one subject improved 7 and estimated that he would lose 2, the difference would be 9; if another subject improved 10 and said he would improve 15, the difference would be 5, and the mean error for the two would be 7. Table 16 presents the percent mean errors for all groups combined.

Size of Error

Though the accuracy of these estimates is not high in comparison with such psychophysical judgments as those of intensity of light, length of line, or even weight, the mean error of the pre-think-estimates (20.1% for cancellation, 16.9% for scribbling, and 22.1% for knot tying) is much better than could be expected from mere wild guessing. The subject does not say, "Oh, I can do twice as well the next time," or anything so wild as that, but on the whole is able, even after only one trial, to judge with some approach to correctness how much he can improve in a second trial.

The accuracy of the may-estimates is inferior to that of think-estimates as might be expected, but only in scribbling and knot tying, not in cancellation. This exception is due to the consistent tendency towards under-estimation in cancellation which is less present in the may-estimates than in the think-estimates.

TABLE 16

Average Percent Mean Error of Estimate for Combined Groups

Trial	*Cancellation*		*Scribbling*		*Knot Tying*	
	%	*S.D.*	*%*	*S.D.*	*%*	*S.D.*
Think Before II	19.3	2.71	14.9	2.70	23.1	4.45
Think Before III	21.0	3.38	18.9	4.29	21.0	2.56
Think Before II & III	20.1	3.18	16.9	2.84	22.1	3.85
May Before II	15.1	2.10	21.1	4.26	30.5	5.23
May Before III	16.8	2.95	26.1	6.92	28.4	5.68
May Before II & III	16.0	2.70	23.6	6.26	29.5	5.56
Think After II	16.7	3.81	15.3	3.17	17.1	1.96
Think After III	18.3	2.87	18.1	2.59	19.2	2.96
Think After II & III	17.5	3.46	16.7	3.26	18.2	2.73
May After II	11.9	2.59	21.3	4.83	24.3	4.65
May After III	14.3	2.96	24.6	4.44	24.0	2.83
May After II & III	13.1	3.16	22.9	4.86	24.1	3.85
Goal	16.7	2.92	25.0	3.74	23.4	4.03
Think After IV	20.1	2.90	20.6	5.26	17.4	2.32
May After IV	14.9	2.59	27.7	5.60	20.7	3.95

Accuracy of First Estimates Compared with That in Following Tasks

Were the first estimates more or less accurate than pre-estimates in later trials? Table 17 presents the average percent difference between the accuracy of pre-estimates in the second and third trials, and in the second and fourth trials. The reliability of the differences is expressed by means of the $S.D._{av.}$, and, as a further check on the smallness of the samples on which the averages are based, the T value is calculated for each difference.

In cancellation and scribbling the accuracy of the pre-estimates is seen to be no greater after two trials than after only one. In fact, it is less accurate (to judge from the minus signs in Table 17) due to the tendency to under-estimate in cancellation and over-estimate in scribbling. In knot tying, the pre-estimates in the second trial are slightly less accurate than those in the third trial.

TABLE 17

Average Percent Difference Between Mean Errors of Pre-Estimates Trial II and Pre-Estimates Trial III and Trial IV

Trial	*Cancellation*			*Scribbling*			*Knot Tying*		
	Av.% Diff.	*S.D.*	$S.D._{av.}$	*Av.% Diff.*	*S.D.*	$S.D._{av.}$	*Av.% Diff.*	*S.D.*	$S.D._{av.}$
Think Before II–	−1.7	2.91	1.10	−4.0	3.63	1.33	2.1	4.19	1.58
Think Before III....			T = 1.20			T = 2.70			T = .92
May Before II–	−1.7	2.19	.83	−5.0	4.60	1.74	2.1	2.70	1.02
May Before III......			T = 1.90			T = 2.66			T = 1.91
Think Before II–	2.6	2.38	.90	−10.1	2.03	.76	−.3	3.33	1.26
Goal...........................			T = 2.67			T = 12.18			T = .22
May Before II–	−1.6	2.44	.92	−4.9	1.46	.55	7.0	4.19	1.58
Goal...........................			T = 1.60			T = 4.88			T = 4.28

The first think-estimates are more accurate than the goal in scribbling and knot tying, but less accurate in cancellation. The reverse is true of the may-estimates.

The differences found between accuracy of pre-estimates in the second trial and accuracy in following trials raise the question as to the manner in which estimates are made without previous knowledge of improvement as compared with methods used when the subjects have had an opportunity to test out improvement. This subject is discussed in detail in Chapter VII.

Accuracy of Pre-Estimates as Compared with That of Post-Estimates

The difference in accuracy between pre- and post-estimates is presented in Table 18. The pre-estimates are less accurate on the whole than post-estimates, but the advantage of the post-estimates is not very great in any one of the tasks. Consider the difference of 2.6% between pre-think-estimates and post-think-estimates in cancellations; or .1% in case of scribbling. This comparative closeness of the pre-estimate accuracy to the post-estimate is further evidence against mere guessing in the pre-estimates; for in the post-estimates the subject has a definite basis for judgment, since he has his actual performance to judge by.

Another conclusion can be drawn from the fact that post-estimates are only moderately superior to the pre-estimates, and that is that the subjects cannot have counted their scores as they made them, *i. e.*, series cancelled, loops made, knots tied, for if they had done so, they need have made no error at all in the post-

TABLE 18

Average Percent Difference Between Mean Errors of Pre- and Post-Estimates

Trial	*Cancellation*			*Scribbling*			*Knot Tying*		
	Av.% Diff.	*S.D.*	$S.D._{av.}$	*Av.% Diff.*	*S.D.*	$S.D._{av.}$	*Av.% Diff.*	*S.D.*	$S.D._{av.}$
Think Before–Think After (II & III)	2.6	2.99	.80 T = 3.13	.1	1.96	.52 T = 1.84	3.9	3.87	1.03 T = 3.63
May Before–May After (II & III)	2.9	2.58	.69 T = 4.06	.7	3.37	.90 T = .75	5.4	5.63	1.24 T = 3.46
Goal–Think After IV	−3.4	.90	.34 T = 9.21	4.4	2.97	1.12 T = 9.59	6.0	3.59	1.35 T = 4.10
Goal–May After IV	1.8	1.46	.55 T = 3.02	−2.7	3.24	1.22 T = 1.54	2.7	2.25	.85 T = 2.94

estimates. Also, if they had counted, their post-may-estimates should have coincided with their post-think-estimates, and they did not. We may conclude, therefore, that this conceivable source of error was not present in the experiment to any serious degree.

Correlation between Estimate of Improvement and Actual Improvement

So far we have been comparing the results for the groups as a whole. The relationship between individual estimates of improvement and actual improvement, *i. e.,* the extent to which estimated improvement can be taken as an indication of actual improvement, remains to be examined. To get at this relationship the correlation between absolute estimates of improvement and actual improvement scores was computed for each of the groups by the Pearson Product-Moment Method. Since the size of the correlation coefficients varied a good deal for the different groups, the difference in the size of the groups was taken into account in averaging the correlation coefficients which appear in Tables 19 and 20. By the use of the z-method, advocated by Fisher for combining values from small samples, a weighted coefficient was found that more nearly represents the average for the group. This coefficient is included in parenthesis after the estimated z-value. (For method employed, see pages 170–171, "Statistical Methods for Research Workers," R. A. Fisher.)

Correlation between Pre-Estimates and Improvement

While a definite relationship between pre-estimates of improvement and actual improvement is shown by the positive correlation in Table 19, the coefficients are not sufficiently high for prognostication. That is, we could not tell with any degree of accuracy where one individual would stand in relation to another in improvement from his estimate of improvement before trial. The correlation increased, on the whole, from trial to trial.

TABLE 19

WEIGHTED AV. CORRELATIONS BETWEEN PRE-ESTIMATES AND IMPROVEMENT

Estimates	*Cancellation*			*Scribbling*			*Knot Tying*		
	z	*S.D.*$_z$	(*r*)	*z*	*S.D.*$_z$	(*r*)	*z*	*S.D.*$_z$	(*r*)
"Think"									
Trial II	.25	.05	(.24)	.16	.05	(.16)	.18	.05	(.18)
Trial III	.37	.05	(.35)	.24	.05	(.24)	.18	.05	(.18)
Goal IV	.51	.05	(.47)	.31	.05	(.30)	.37	.05	(.35)
"May"									
Trial II	.32	.05	(.31)	.12	.05	(.12)	.19	.05	(.19)
Trial III	.37	.05	(.35)	.25	.05	(.24)	.35	.06	(.34)
Trial IV	....	....	....	....	....	....	....	....	....

Correlation between Post-Estimates and Improvement

When we examine the correlation coefficients between post-estimates of improvement and actual improvement we find a somewhat more significant relationship than was found with the pre-estimates. The coefficients were also found to be more uniform for the different groups. The weighted average correlations calculated by the z-method are as follows:

TABLE 20

WEIGHTED AV. CORRELATIONS BETWEEN POST-ESTIMATES AND IMPROVEMENT

Estimates	*Cancellation*			*Scribbling*			*Knot Tying*		
	z	*S.D.*$_z$	(*r*)	*z*	*S.D.*$_z$	(*r*)	*z*	*S.D.*$_z$	(*r*)
"Think"									
Trial II	.61	.05	(.54)	.39	.05	(.37)	.73	.05	(.62)
Trial III	.72	.05	(.62)	.29	.05	(.28)	.63	.05	(.56)
Goal	.69	.05	(.60)	.32	.05	(.31)	.61	.05	(.54)
"May"									
Trial II	.69	.05	(.60)	.34	.05	(.33)	.71	.05	(.61)
Trial III	.53	.05	(.49)	.35	.05	(.34)	.71	.05	(.61)
Trial IV	.76	.05	(.64)	.34	.05	(.33)	.61	.05	(.54)

The difference between the relationship here and that found in the pre-estimates can be determined by subtracting the correlation coefficients for each group, and averaging these differences. By this method we find in cancellation that the post-estimate correlations are .20 (S.D. .19, S.D.$_{av.}$ .04) higher than the pre-estimates for the think-estimates and .24 (S.D. .19, S.D.$_{av.}$ .05) for the may-estimates, which indicates a reliable difference. In scribbling the difference is not so marked: .09 (S.D. .19, S.D.$_{av.}$ .04) for think-estimates, .16 (S.D. .16, S.D.$_{av.}$ .04) for may-estimates. It is to be noted that the correlations themselves are lower for this task. In knot tying the difference is greatest of all: .27 (S.D. .22, S.D.$_{av.}$ .05) for think-estimates, and .32 (S.D. .20, S.D.$_{av.}$ .05 for may-estimates.

We do not find the increase in the relationship between post-estimates and improvement from trial to trial that we found between pre-estimates and improvement. However, it is to be remembered that the subjects were not informed of the accuracy of their estimates of improvement. If they had been so informed they might have gained in accuracy and shown an increasingly high relationship between post-estimates and actual improvement.

We have to conclude in regard to the post-estimates, as well as the pre-estimates, that while the accuracy of these estimates is fairly great, the individual estimates do not correspond closely enough with actual improvement to enable one to judge the relative position of an individual as regards his improvement from his estimate of improvement either before or after trial.

The correlations between estimated improvement and actual improvement for the may-estimates do not vary greatly nor consistently from the correlations for the think-estimates.

VI. INDIVIDUAL CONSISTENCY IN ESTIMATES

While the subjects' estimates are not reliable indicators of relative position in improvement, they may show consistency for a given individual. That is, we may find the subjects falling into such classifications as "high" or "low" estimators; "good" or "poor" estimators; "over"- or "under"-estimators, regardless of the task to be performed. Or we may find a good deal of individual consistency throughout a given task.

INDIVIDUAL CONSISTENCY IN PERFORMANCE OF DIFFERENT TASKS

Before examining the consistency in estimates themselves, we must first get an idea of the relationship between performance in the three tasks. In Table 21 we have the weighted average correlations between original scores and between absolute improvement scores in the three tasks. These average coefficients show a low positive correlation between performance in the three tasks, the greatest consistency being found between cancellation and scribbling.

TABLE 21

WEIGHTED AV. CORRELATIONS BETWEEN PERFORMANCE IN THE DIFFERENT TASKS

	Can. & Scrib.			*Can. & Knots*			*Scrib. & Knots*		
	z	$S.D._z$	(r)	z	$S.D._z$	(r)	z	$S.D._z$	(r)
Original Score	.30	.05	(.29)	.18	.05	(.18)	.08	.05	(.08)
Improvement II	.07	.05	(.07)	.03	.05	(.03)	.07	.05	(.07)
Improvement III	.24	.05	(.24)	.02	.05	(.02)	.11	.05	(.11)
Improvement IV	.26	.05	(.25)	.06	.05	(.06)	.05	.05	(.05)

From the low correlations we can infer that the tasks call out fairly distinctive types of behavior. If the estimates show definitely higher correlations than shown between performance, it would be reasonable to infer the presence of a general estimating behavior trait. If, on the other hand, the correlations are not noticeably higher, it would seem that the estimating process is specific to the task involved.

INDIVIDUAL CONSISTENCY IN SIZE OF ESTIMATES

We judge whether an individual's estimate of improvement is "high" or "low" in terms of the absolute estimates of his fellow subjects, or in terms of percent of the original score, not in terms of actual improvement. In scribbling, for instance, an estimated improvement of 20 would be considered high, no matter what the actual improvement, since the average for the groups as a whole is below 10; likewise an estimated improvement of 49% in this task would be considered high, even if the individual actually improved more than this, since the average for the groups as a whole, for any trial, does not exceed 34%.

A tendency toward high or low estimation would seem to be a broader and simpler behavior trait than accuracy of estimation, which requires a fine discrimination in differences in performance. For this reason we might expect to find some consistency between size of estimates in different tasks, even though consistency between performance in these tasks or consistency in accuracy might be lacking.

Two questions to be answered in considering this problem of individual consistency in size of estimates are: Do the estimates show some agreement as between tasks? and—Are the correlation coefficients between the size of estimates in the three tasks sufficiently high to warrant a classification of ''high'' or ''low'' estimators, irrespective of the task to be performed?

The correlation coefficients between absolute estimates of improvement for the different tasks, presented in Table 22–A, indicate that individuals do show some tendency to high or low estimates—some consistency in this respect—without regard to their consistency in performance, or to accuracy of estimation. However, since the majority of the correlation coefficients are well below 40, we must assume that the variation in the kinds of activity called forth by the different tasks modified the estimating process in the different tasks, thus obscuring any marked tendency towards high or low estimates.

Individual Consistency in Accuracy of Estimates

As a rule, in order to estimate accurately a subject had to consider the size of his original score carefully, since close approximation between estimates and improvement would not be brought about frequently by mere chance. He had also to be alert to the various cues afforded him by which he might judge his performance. In scribbling, for example, a subject with a high original score would have to realize that his chances for improvement were limited. The process of estimating accurately is fairly complex, since it takes into account a number of different elements. For this reason it might not show up so clearly as the simpler tendency towards high or low estimation. By looking at Table 22–B, in which the coefficients of correlation between absolute errors of estimate for the different tasks are given, we find this to be the case. Not only are the coefficients lower than those found between size of estimates, but also, in some instances a slight negative relationship is found.

TABLE 22

Weighted Av. Correlations Between Estimates of Improvement in Different Tasks

Estimates	*Can. & Scrib.*			*Can. & Knots*			*Scrib. & Knots*		
	z	*$S.D._z$*	*(r)*	*z*	*$S.D._z$*	*(r)*	*z*	*$S.D._z$*	*(r)*
A. Correlations Between Absolute Estimate of Improvement									
Think Before II	.23	.05	(.23)	.08	.05	(.08)	.28	.05	(.27)
Think Before III	.43	.05	(.41)	.13	.05	(.13)	.25	.05	(.24)
May Before II	.40	.06	(.38)	.23	.06	(.23)	.42	.06	(.40)
May Before III	.60	.05	(.54)	.24	.06	(.24)	.21	.06	(.21)
Think After II	.11	.05	(.11)	.10	.05	(.10)	.10	.05	(.10)
Think After III	.38	.05	(.36)	.19	.05	(.19)	.23	.05	(.23)
May After II	.11	.05	(.11)	.06	.06	(.06)	.20	.05	(.20)
May After III	.38	.05	(.36)	.20	.06	(.20)	.25	.06	(.24)
Goal	.45	.05	(.42)	.27	.05	(.26)	.34	.05	(.33)
Think After IV	.38	.05	(.36)	.17	.05	(.17)	.24	.05	(.24)
May After IV	.33	.05	(.31)	.20	.05	(.20)	.28	.05	(.27)
B. Correlations Between Size of Errors of Estimation (Accuracy)									
Think Before II	.01	.05	(.01)	−.04	.05	−(.04)	.02	.05	(.02)
Think Before III	.06	.05	(.06)	.01	.05	(.01)	.02	.05	(.02)
May Before II	−.07	.06	−(.07)	.02	.06	(.02)	.11	.06	(.11)
May Before III	.02	.05	(.02)	−.04	.06	−(.04)	.16	.06	(.16)
Think After II	.03	.05	(.03)	.04	.05	(.04)	−.07	.05	−(.07)
Think After III	.09	.05	(.09)	.19	.05	(.19)	.08	.05	(.08)
May After II	−.05	.05	−(.05)	.05	.06	(.05)	.08	.05	(.08)
May After III	−.01	.05	−(.01)	.12	.06	(.12)	.07	.06	(.07)
Goal	−.03	.05	−(.03)	.01	.05	(.01)	.03	.05	(.03)
Think After IV	.07	.05	(.07)	.04	.05	(.04)	−.07	.05	−(.07)
May After IV	−.03	.05	−(.03)	.00	.05	(.00)	.11	.05	(.11)
C. Correlations Between Over- and Under-Estimation (Constant Error)									
Think Before II	.05	.05	(.05)	.07	.05	(.07)	.18	.05	(.18)
Think Before III	.15	.05	(.15)	.10	.05	(.10)	.19	.05	(.19)
May Before II	.00	.06	(.00)	.12	.06	(.12)	.17	.06	(.17)
May Before III	.21	.05	(.21)	.16	.06	(.16)	.25	.06	(.24)
Think After II	.12	.05	(.12)	.00	.05	(.00)	.17	.05	(.17)
Think After III	.12	.05	(.12)	.16	.05	(.16)	.12	.05	(.12)
May After II	.13	.05	(.13)	.08	.06	(.08)	.14	.05	(.14)
May After III	.19	.05	(.19)	.20	.06	(.20)	.17	.06	(.17)
Goal	.14	.05	(.14)	.15	.05	(.15)	.08	.05	(.08)
Think After IV	.16	.05	(.16)	.18	.05	(.18)	.06	.05	(.06)
May After IV	.23	.05	(.23)	.22	.05	(.22)	.13	.05	(.13)

This disagreement in accuracy in cancellation and scribbling, the two tasks that showed the most agreement in performance, suggests that the estimating process in the two tasks was quite different.

Individual Consistency in Over- or Under-Estimation

Granting that individuals do not show marked consistency in accuracy of estimates of improvement when the size of error alone is considered, may there not still be a distinction between those who tend to over-estimate or under-estimate, irrespective of the task required?

That there is some tendency towards consistency in over- and under-estimation in the three tasks is evident from the positive correlation coefficients between individual errors in the three tasks found in Table 22–C. The sign of the error, as an over- or under-estimate, is here taken into account. The correlation coefficients are not sufficiently high, however, to warrant the segregation of a possible general factor by means of the triad difference formula. In fact, it is probable that the direction of error (over- or under-estimation) was a by-product of the experiment—a result of the arithmetical calculation in which the estimate score appeared as either higher or lower than the performance score; rather than an individual behavior trait, since the subjects did not have any way of checking up on the accuracy of their estimates, and thus of getting an idea of actual performance. If they had known in each trial what their previous performance had been, and had then consistently under- or over-estimated throughout the three tasks, we might have had grounds for postulating the presence of such a general factor.

In summing up the data on consistency in estimates as between the three tasks we conclude that in this experiment the classifications of ''high'' or ''low'' estimators, ''good'' or ''poor'' estimators, ''over- or under-estimators'' are not warranted. The fact that consistency is somewhat greater in size of estimates than in accuracy or in over- and under-estimation, indicates a tendency to high or low estimates, which is obscured when actual improvement is taken into account.

Individual Consistency of Estimates Throughout the Same Task

So far we have been considering individual consistency in estimating in different tasks. How about consistency in a given task? Is there a close relationship between pre- and post-estimates in the same trial, or between post-estimates in one trial and pre-estimates in the following trial?

Correlation between Pre- and Post-Estimates

That there is a definite relationship between pre-estimates and post-estimates is shown by the fairly high weighted average correlation coefficients between absolute pre- and post-estimates for the different trials, presented in Table 23. From these positive correlations between pre- and post-estimates we can infer that the subjects were influenced in making their estimates after performance by what they had estimated their improvement would be before trial. Later reports by the subjects verified this supposition. The fact that the coefficients in the third and fourth trials are higher than those in the second trial suggests that the tendency to base post-estimates on pre-estimates becomes established with practice.

TABLE 23

WEIGHTED AV. CORRELATIONS BETWEEN PRE- AND POST-ESTIMATES

Estimates	*Cancellation*			*Scribbling*			*Knot Tying*		
	z	*S.D.*$_z$	(*r*)	*z*	*S.D.*$_z$	(*r*)	*z*	*S.D.*$_z$	(*r*)
"Think"									
Trial II	.43	.05	(.41)	.53	.05	(.49)	.44	.05	(.41)
Trial III	.74	.05	(.63)	.83	.05	(.68)	.53	.05	(.49)
Goal	.71	.05	(.61)	.70	.05	(.60)	.57	.05	(.52)
"May"									
Trial II	.54	.05	(.49)	.50	.05	(.46)	.44	.06	(.41)
Trial III	.76	.05	(.64)	.83	.05	(.68)	.59	.06	(.53)
Trial IV	.78	.05	(.65)	.76	.05	(.64)	.58	.05	(.52)

Correlation between Estimates in Different Trials

In making their pre-estimates in the third trial, were the subjects influenced by the estimates they had just made after the second trial? The fairly high weighted average correlations found in Table 24, between absolute post-estimates in the second trial and pre-estimates in the third trial indicate that such was the case, although subjects reported later they made their estimates independently in each trial. The correlations between absolute post-estimates in the third trial and the goal, also given in Table 24, bear out the supposition that some sort of relationship is present. It will be noticed that these correlations approximate in size the correlations between pre- and post-estimates in Table 23.

The correlations between these estimates might be explained by the fact that the subjects established a certain mental set in estimating. That is, in their first pre-estimates they probably selected an amount of improvement which served as a base line for their

TABLE 24

WEIGHTED AV. CORRELATIONS BETWEEN POST-ESTIMATES IN ONE TRIAL AND PRE-ESTIMATES IN THE FOLLOWING TRIAL

Estimates	*Cancellation*			*Scribbling*			*Knot Tying*		
	z	$S.D._z$	(r)	z	$S.D._z$	(r)	z	$S.D._z$	(r)
Post-Est. II & Pre-Est. III									
“Think”	.68	.05	(.59)	.65	.05	(.57)	.60	.05	(.54)
“May”	.64	.05	(.56)	.69	.05	(.60)	.62	.06	(.55)
Post-Est. III & Goal									
“Think”	.66	.05	(.58)	.65	.05	(.57)	.47	.05	(.44)
“May”	.81	.05	(.67)	.78	.05	(.65)	.49	.05	(.45)

later estimates, although they may have done this without consciously making use of their estimates. This theory receives support from the greater consistency between post-estimates of improvement in one trial and pre-estimates in the following trial, than is found between actual improvement in one trial and pre-estimates in the following trial. (See Table 25.)

TABLE 25

WEIGHTED AV. CORRELATIONS BETWEEN IMPROVEMENT IN ONE TRIAL AND PRE-ESTIMATES IN THE FOLLOWING TRIAL

Estimates & Imp.	*Cancellation*			*Scribbling*			*Knot Tying*		
	z	$S.D._z$	(r)	z	$S.D._z$	(r)	z	$S.D._z$	(r)
Imp. II & Est. III									
“Think”	.37	.05	(.35)	.24	.05	(.24)	.33	.05	(.31)
“May”	.37	.05	(.35)	.23	.05	(.23)	.27	.06	(.26)
Imp. III & Goal	.55	.05	(.50)	.21	.05	(.21)	.41	.05	(.40)

In other words, in estimating improvement before the third and fourth trials, the subjects seemed to have been influenced more by what they had estimated for the preceding trial than by what they had actually accomplished. That the correlations between post-estimates in the second trial and pre-estimates in the third trial are reliably higher than the correlations between improvement in the second trial and pre-estimates in the third trial is shown by the average differences between the two. The average differences for the third and fourth trials are not so reliable. (See Table 26.)

In summarizing the results obtained on individual consistency in estimating throughout a given task, we can say that the subjects showed definite consistency in their pre- and post-estimates in the

TABLE 26

AVERAGE DIFFERENCE BETWEEN CORRELATIONS BETWEEN (A) POST-ESTIMATES IN ONE TRIAL AND PRE-ESTIMATES IN THE FOLLOWING TRIAL; AND (B) IMPROVEMENT IN ONE TRIAL AND PRE-ESTIMATES IN THE FOLLOWING TRIAL

		Trials II and III (*A–B*)			*Trials III and IV* (*A–B*)		
		Av. diff.	*S.D.*	$S.D._{av.}$	*Av. diff.*	*S.D.*	$S.D._{av.}$
Can	Think	.16	.16	.06	.11	.24	.09
	May	.27	.13	.05	.15	.29	.11
Scrib	Think	.32	.21	.08	.38	.19	.07
	May	.38	.22	.08	.42	.16	.06
Knots	Think	.21	.09	.03	.03	.22	.08
	May	.30	.13	.05	.04	.25	.09

same trial, as well as consistency in post-estimates in one trial and pre-estimates in the following trial. The fact that the subjects were not given opportunity to check up on the accuracy of their estimates in the course of the experiment may have been partly responsible for this consistency between estimates in the same task.

GROUP, AGE AND SEX DIFFERENCES

While the subjects utilized were all adults, the make-up of the seven groups differed considerably. Hence there was a possibility that some group consistency in estimating might appear. A careful examination of averages and correlations, however, failed to bring out a significant difference between the groups.

The inclusion of the high school group introduced an age factor that might have been operative in distinguishing this group from the rest. From the fact that such a distinction was not found, we conclude that either the difference in age between this group and the others was not great enough to cause an appreciable difference in results, or that the age factor was relatively unimportant in this experiment.

The Science and Industry groups were composed entirely of men; the Margaret Morrison group entirely of women. The Pitt., Arts and high school groups included a fair proportion of both men and women. By comparing the results of the men and women it was possible to determine the significance of the sex factor in estimating. Again, the results failed to show any marked difference between the groups.

VII. INTERPRETATION OF OUTSTANDING RESULTS

The main question to be answered by an experiment of this sort concerns the process of estimating and the factors that determine the estimate. Various indications from the quantitative results bear upon this problem, and will be brought up from the preceding sections and incorporated here. A number of questions suggested themselves during the study of the quantitative data which could only be answered by a more minute observation of the individual's behavior during the experiment, aided by his own report of his procedure. A questionnaire was therefore prepared which served to bring out further information on what was taking place during the estimating process. This questionnaire was filled out immediately after the experiment by the fifteen graduate students who took part in an investigation on the goal to be reported in this section. Ten subjects furnished additional information on how they made their estimates the first time, since the questionnaire did not bring out sufficient material on this point.

Questionnaire Utilized in Securing Introspective Data

1. Did you actually refer to your original score each time in making your ''think'' estimate before performing the task? If not, what did you do?
2. In making your ''think'' estimate after performance, did you base this estimate on the original score or on the before estimate for that trial?
3. In making your ''think'' estimate, did you consciously make use of your estimate of improvement in the preceding trial, or did you make a relatively independent estimate in each trial?
4. Did you actually calculate improvement in most cases or did you merely guess?..In which task was there the least amount of guessing?..In which the greatest amount of guessing?..
5. Did you base your ''may'' estimates on the original score or on the ''think'' estimates?
6. In making your ''may'' estimates, which of the following methods did you use? Check the one (or ones) that applies.
    Attempted to make a cautious estimate of the greatest possible improvement.
    Simply added a few more units to the ''think'' estimates.
    Took this occasion to let myself go.
7. Did you consciously change your method of procedure in any task? If so, how? Did this affect your estimates in any way?
8. Did you change your method of estimating during the experiment? If so, in what tasks, and how?
9. What specific cues did you use in estimating performance?
    For cancellation ....
    For scribbling ....
    For knot tying ....
10. In which task do you think you estimated improvement the most accurately?..................................In which the least accurately?..................................
11. Which task did you think was the hardest to do?..
    Which was the easiest to do?..

12. Which task did you like the best?..Which task did you like the least?..Did you dislike any of the tasks? If so, why?

DIFFERENCES IN THE THREE TASKS WITH RESPECT TO ACTUAL AND ESTIMATED IMPROVEMENT

One of the outstanding results of the quantitative study is the difference between the three tasks with respect to the amount of actual improvement, and with respect to estimated improvement. It will be remembered that actual improvement is greatest in cancellation and least in scribbling, whereas high estimates of improvement predominate in scribbling and low estimates in cancellation. Knot tying shows no tendency in either direction.

A consideration of the comparative difficulty of the three tasks affords an explanation of the above paradox. In making this comparison the point is brought home that the ease of a task is not synonymous with improvability in the task, although it is probable that the two concepts were confused by the subjects.

Scribbling is undoubtedly the easiest of the three tasks, and was reported so by the subjects. It involves a fairly automatic type of response and not much skill. This task brought forth a great display of energy on the part of the subjects. They went at it with a ''do-or-die'' attitude and in their enthusiasm tended to bear down hard on the paper and to get into tense positions that did not necessarily result in improved performance. The subjects were frequently misled into believing that this excess energy would result in increased output. As an actual fact, scribbling is closely akin to tapping in which subjects tend to reach their physiological limit quickly. It will be recalled that there was no appreciable increase in improvement in this task after the second trial. The fact, then, that scribbling seemed easy, but actually did not lend itself to much improvement is responsible in part for the subjects' overestimation in this task.

Cancellation, on the other hand, was felt by the subjects to be more difficult. Furthermore, this task was taken most seriously, the subjects showing close concentration on the cancellation sheet. The procedure in this task is pretty well routinized. In order to scrutinize the figures carefully, a limited rate of speed must be maintained and the figures examined in an orderly fashion. This curtailment of variation in procedure tended to make performance in the different trials seem more or less alike to the subjects, and hence conceal the presence of improvement. Cancellation brings into play higher perceptual reactions than those called out in scribbling, and allows for a good deal more improvement. The fact that the subjects recognized this as more of a ''mental test'' than scribbling probably accounted somewhat for their under-estimation in this task. In spite of their under-estimation, the subjects liked this task the best.

Knot trying was considered the most difficult task, and liked the least. The subjects showed quite a bit of nervousness throughout the knot tying, invariably starting with a certain strained amusement, followed by a rather grim readiness. They became quite verbal in this task, displaying more emotion than in the case of scribbling or cancellation. The reasons for disliking this task were also of emotional origin. In two cases (out of the 15 special subjects) of actual dislike, one subject gave as a reason the fact that her fingers shook and it made her mad, the other that it made her feel awkward. Impatience and irritation over difficulties with the knots lessened the subjects' confidence in improvement and caused variations in the estimating process. Sometimes a sense of failure at the beginning of this task lasted throughout the experiment. This task called for quite a bit of skill in manipulation, and losses as well as gains in performance were frequent. As a result, we get a mixture of over- and under-estimation on the part of the subjects.

The subjects reported the least amount of guessing in cancellation, and the most in scribbling. They stated that their estimates on the whole were in the nature of a guess, which indicates a lack of assurance in their estimates.

As we study the three types of tasks, knot tying stands out as the type of performance more nearly allied to every day experience. There is room for a certain amount of improvement in this task, complicated to be sure by variations in performance, and this improvement is more or less self-evident. In scribbling improvement is restricted, but seems possible; in cancellation improvement is possible but seems limited. As a result we get a balance between over- and under-estimation in the first, over-estimation in the second; and under-estimation in the third.

INITIAL ESTIMATES

One of the most interesting questions is as to how the subjects estimated improvement the first time, before they had repeated the task. Why, for example, without previous experience with improvement, did they under-estimate in cancellation and over-estimate in scribbling? The response to the questionnaire did not satisfactorily answer these questions, since the cues mentioned as an aid in estimating were, for the most part, those which might come through practice.

In order to get at the problem, a modification of the experiment was performed with ten adult subjects. They were given the three tasks, in a group, in the prescribed manner, and the tasks were then scored. They were next told that they would be asked to repeat each of the tasks once. Instructions for making estimates of improvement were then given (using the standard directions), but this time the estimates for the three tasks were made in direct succession, without the trial intervening. This was done to avoid later confusion with post-estimates. After the estimates were made (only the top third of the estimate blank was used—see Figure 1) the following instructions were given: ''Before starting the second trial, turn to the blue sheet (cancellation estimate blank) once more. On the sheet of paper attached, I want you to tell me as simply as possible just how you made your estimate. In other words, why did you estimate that particular amount rather than any other?—Turn to the tan sheet and describe in the same way just how you made the estimate in this case.—Now do the same for knot tying on the sheet attached to the pink blank.'' At the request of the subjects the tasks were then repeated.

In cancellation, the general feeling that with practice one should be able to improve any performance by a moderate amount guided most of the subjects in their estimates. Familiarity with the cancellation blank and with the experiment in general was given as another basis for estimated improvement. In two instances estimated improvement was based on the determination to avoid the errors on the first trial. (This was an erroneous basis for estimation since the identical blank was not used the second time.) In one instance the subject estimated that her score would remain the same because she felt her rate of speed had been as rapid as she could make it, and that greater speed would only make more errors. (This subject improved 35% in the second trial!) On the whole, then, in cancellation a moderate estimate of improvement was made on the assumption that practice should produce some improvement. No mention was made of better methods of procedure, other than the avoidance of errors.

In scribbling the scoring of the task was an essential factor in estimating improvement. Through scoring the subjects became conscious of the type of scribble that would produce the highest score. Improvement in the type of the scribble was always anticipated.

In knot tying also change in method accounted for estimates of improvement. Such comments as, "My knots all went in the same place the first time, but I may be able to do a few more next time" (performance four times greater in second trial!). "I know that it is necessary not to hurry so that all motions are smooth," and "I will not tie any knots on top of each other," illustrate the type of change in method indicated. The emotional factor entered in the two cases in which the subjects estimated that the score would remain the same, as seen in the remarks, "I'd never do any better because I get the jitters with my hands," and "I had a feeling I did my best the first time—I may get excited and not do so well."

First estimates on the whole, then, were based on a general belief in ability to improve with practice, plus anticipation of improved methods of performance in scribbling and knot tying. The subjects were not able to tell the reason for the exact amount of estimate other than that they felt a moderate amount of improvement could be made. While these subjects did not mention size of the original score in explaining their basis for estimating, that the original score was taken into account in making the estimates has been shown by the fact that the percent estimates of those with the highest original scores in the Pitt. group were decidedly lower than the percent of those with the lowest scores. That this was not taken into account sufficiently to avoid error, was also shown by the fact that in the main experiment those with the highest original scores over-estimated more or under-estimated less than those with the lowest scores. This was true not only of the pre-estimates in the second trial, but in the third and fourth trial as well.

PRE- AND POST-ESTIMATES

How do subjects estimate improvement after the task has been repeated once? Do the pre-estimates influence the post-estimates? Do the post-estimates in one trial influence the pre-estimates in the following trial?

Some information on the above questions is supplied by the quantitative data already discussed. We found, for example, that the post-estimates in the second and third trials were slightly higher than the pre-estimates in cancellation, and somewhat lower in scribbling and knot tying. This resulted in a very slight superiority in percent of accuracy in post-estimates over pre-estimates as well as a smaller percent of under-estimation in cancellation and over-estimation in scribbling as compared with pre-estimates. The fact that the differences between the pre- and post-estimates were so small, argues against wild guessing in the pre-estimates, and against actual knowledge of improvement, such as might come from counting the knots the scribbles made, or the lines covered in cancellation, in the post-estimates. We can also infer from the small difference between pre- and post-estimates a close similarity in the basis of estimation in the two cases.

The relationship between pre- and post-estimates was brought out statistically in the correlations between them and between post-estimates in one trial and pre-estimates in the following trial. We found a fairly high degree of correlation between pre- and post-estimates. The subjects seemed to be guided, then, in setting up their post-estimates, by what they had estimated before trial, although at the same time they tended to correct for over- or under-estimation in pre-estimates. The presence of a definite relationship between post-estimates of improvement in one trial and pre-estimates in the following trial indicated either that they had a general set in the size of estimates, which brought about this relationship, or that they actually based their pre-estimates on their post-estimates.

Let us consider what the subjects reported as to the influence of one estimate on another. The response to the questionnaire indicated that they used the original score in making their pre-estimates but based their estimates after trial more frequently on the pre-estimates than on the original score. So far, the introspective account agrees with the quantitative findings. When we come to the question of the influence of estimates from trial to trial, we find the subjects reporting that their estimates as a rule were made independently in each trial. The correlations between post-estimates in one trial and pre-estimates in the following trial do not substantiate this report. While the subjects may have not consciously utilized the estimates in one trial in estimating in the succeeding trial, there was an influence present, whether the subject recognized it or not.

What else guided the subjects in making their post-estimates? The main objective cue in cancellation was the position on the sheet that the subject had reached when time was called. This indicated the amount covered but not the accuracy. In scribbling the character of the loops made, large or small, sharp or round, the general appearance of the page, distance between loops, number of columns made, and tenseness of muscle, were given as indications of improvement. These cues were somewhat more definite than the general position on the page, reported in cancellation. At the same time, it was not easy to estimate the number of loops on a page by merely glancing at it, even though the subject was aware that the general nature of the performance had improved. Knot tying afforded the most definite cues, such as the number of double knots made (which did not count as two), the smaller amount of space between knots, the ''feel'' of the knots as they slipped through the fingers, the ease with which the fingers manipulated the knots, and actually counting the knots contrary to instructions. While offhand it may seem that the possibility of counting the knots was an important factor in directing post-estimates, it did not enter into the situation to a great extent, as has been mentioned, or the post-estimates would have been more accurate. Furthermore, the subjects were so intent on tying the knots that they had little time left for counting. The tendency to count came on the completion of the experiment, and the investigator attempted to avoid this contingency by ordering the string to be put back in the envelope immediately.

Changes of performance during the experiment were mentioned for scribbling and knot tying, but not for cancellation. In scribbling the changes consisted in making smaller loops, holding the paper differently, pulling the paper under the pencil, changing arm position, making certain that the first mark would be a loop, and ''scribbling faster.'' By faster, in this case the subject probably meant harder as well. The following changes were reported in tying knots: starting at the center of the string to avoid overlapping knots, making looser and different types of knots, turning the string around and starting from both ends, holding the knot in place in pulling the string through, working clear of the table, removing finger ring, trying to work more carefully. The subjects reported that these changes in method of performance raised their estimates of improvement, but did not change their method of making the estimates. This does not mean that there was an increase in accuracy of estimates with the change in method, for it will be recalled, we do not find an increase in accuracy of estimates in any of the tasks from trial to trial.

In order to secure more detailed information regarding the basis of estimates than that afforded by a statement of cues used by the subjects, a scheme of introspection was prepared, based on that employed by Hollingworth in his study, ''Perceptual Criteria of Judgments of Efficiency,''* and included as part of the questionnaire.

---

* Experimental Studies in Judgment, Chapter II, pages 18, 19.

Basis for Judgment of Improvement

Think over the way in which you judged improvement in performance in each of the tasks. Arrange the factors given below in the order of their importance with respect to the degree to which they constituted the basis or criteria for your judgments of improvement. Place the most important first, then the next in importance, etc., using the letter for identification. Do this separately for each of the three tasks.

A. Feelings of ease or comfort, or of strain and uncertainty, as the task proceeds.
B. Feelings of pleasantness and satisfaction, or of unpleasantness and dissatisfaction, either during the task or after its completion.
C. The perception of the smoothness and regular flow or of the roughness and irregularity of the performance.
D. Direct estimate of the number of units completed as compared with the original score, regardless of what happens during the performance of the task.
E. Perception of the speed or rate of succession of the separate acts which the task involves.
F. Inference based on the number or amount of specific mistakes, hesitations, changes of method, during the task or remembered after its completion.
G. Feelings of surprise or of fulfilled or unfulfilled expectation when the end of the task is reached.
H. Unanalyzable and indefinite feeling of efficiency or of inefficiency.
I. Any other specific criteria which you may have noted. (Indicate here ....)

Cancellation	*Scribbling*	*Knot Tying*
1—	1—	1—
2—	2—	2—
3—	3—	3—
4—	4—	4—
5—	5—	5—
6—	6—	6—
7—	7—	7—
8—	8—	8—
9—	9—	9—

The subjects reported the use of all eight criteria for estimating improvement. One subject added the feeling of definite failure in knot tying as a criterion. The final position of importance for each criterion determined by averaging the fifteen arrangements for the given task is as follows:

	Cancellation	*Scribbling*	*Knot Tying*
A	3.6	3.5	4.1
B	4.6	3.5	3.6
C	3.9	2.9	3.1
D	3.3	5.7	5.4
E	3.3	2.5	4.5
F	5.9	6.5	4.0
G	5.8	5.3	6.0
H	5.8	6.2	5.5
I	..	..	..

In Hollingworth's experiment the subjects were estimating performance in four repeated tests, namely, color naming, opposites, cancellation and addition, in terms of "better than usual" and "worse than usual." Since the tests were given by the amount limit method, judgments were made on the basis of time taken to complete a given unit of work rather than on amount of work completed in a given time as was the case in the present experiment. In spite of this dissimilarity in procedure the final order for the different criteria used in the two experiments showed agreement in a number of respects.

FINAL ORDER FOR CRITERIA

Present Experiment	C-E-A-B-D-F-G-H
Hollingsworth's Experiment	E-C-F-A-B-H-D-G

In both experiments speed of separate acts (E) and smoothness of performance (C) held a high place. Feelings of surprise (G) and indefinable feelings (H) did not play an important rôle.

Inference based on mistakes and changes of method (F) occupied a lower place in the present experiment than in Hollingworth's. This result seems to agree with the fact discovered that although the subjects reported changes in method used, such changes had only the general effect of raising the estimates. Direct estimate, independent of what occurs during the performance, was rated higher in the present experiment than in Hollingworth's. This difference may be due to the fact that the use of the original score as a basis for estimate emphasized the independent estimate, or because the subjects were estimating in terms of amount of work completed and not in time taken to complete the task.

In comparing the criteria used in the three tasks, we find that in knot tying change in method was more important as a criterion than in the other two tasks, while perception of speed or rate of separate acts was utilized less frequently. Estimation in cancellation depended to a greater extent on the direct estimate of the number of units completed as compared with the original score, regardless of what happened during the performance. This is easily understood since very little actually happened during performance in this task, to judge from the fact that no changes were reported in methods used.

Certain characteristic difficulties in the tasks have been mentioned to account for the differences in the estimates. Is there anything in the nature of the cues utilized in estimating in the three tasks that might also help to explain these differences? We are impressed at once with the fact that there were fewer cues provided in cancellation than in the other two tasks. Furthermore, the lack of changes in performance in this task during the progress of the experiment did away with the tendency found in the other two tasks to raise the estimate as the result of a change in method. These two factors undoubtedly had their part in causing the subjects to under-estimate in this task.

Scribbling and knot tying, on the other hand, involved motor responses in which variations in performance were more easily discernible than in the perceptual responses called out in cancellation. As a result, we find the subjects reacted to indications of improvement more readily in these two tasks. These results are in agreement with the conclusions arrived at by Hollingworth in a study of confidence in judgments, in which he states:*

"Progressively larger variations in performance are required as the basis for judgments of a given degree of confidence as one passes from an automatic, objectively observable, motor performance (such as tapping) through work involving perceptional reactions (color-naming) to work of a more strictly mental and less objectively observable character (opposites)."

While there were definite cues which guided the subjects in estimating improvement in scribbling, they were less easily discernible than those in knot tying, and consequently less active as a check on over-estimation. It was also easier for the subjects to see the relationship between changes of method of performance and improvement in knot tying than in scribbling. In scribbling, for instance, a subject might know that by changing from a large scrawl to a smaller one he probably improved over his original score, and yet over-estimate this improvement quite a bit. Whereas, in knot tying, a subject could estimate pretty closely the precise effect of a change in method that did away with over-lapping of knots. As a result, we find the tendency towards over- and under-estimation balancing each other in knot tying.

The similarity between the cues used by the subjects in making their first estimates and in making their later estimates helps to explain why in cancellation and scribbling the first estimates were more accurate than the pre-estimates in the third trial. (Tables 13 and 17.) In cancellation, it will be recalled, they based their estimates of improvement the first time on the assumption that a moderate amount of improvement should be possible through practice. They under-estimated the amount, however. Not having any definite check on the accuracy of estimates other than the few cues afforded during performance (and the difference between the accuracy of pre- and post-estimates was slight) they persisted in estimating moderate improvement, while they actually improved considerably. In scribbling, on the other hand, first estimates of improvement were based largely on proposed change in method, or familiarity with the task. That is, they felt that they had ''gotten the hang of it'' in the first performance and could do better the second time. They over-estimated this factor at the start. As a result, they continued estimating too much improvement throughout the experiment. Since they actually improved practically not at all after the second trial, their accuracy of estimation decreased. The tendency to confuse excess energy with increased output only made matters worse. In this case the cue was a hindrance rather than a help. That this was an important factor is borne out by the high place rate of speed of separate acts, held as a criterion for basis of estimation. As has already been intimated, in knot tying the specific nature of the cues helped the students to correct errors of estimation. As a result we find the estimates in the third trial more accurate than those in the second, when size of error alone is considered (Table 17), and over-estimation in the pre-think estimates in the second trial changing to under-estimation in the third trial (Table 13). Changes in method of tying knots held a high place as a criterion for basis of estimation right from the beginning of the experiment.

---

* Hollingworth, Experimental Studies in Judgment, Chapter I, page 16.

The fact that the subjects tended to base their post-estimates on their pre-estimates destroyed a good deal of the advantage that the cues afforded, and is a definite explanation of the relatively small difference between the pre- and post-estimates, as well as the reason why there was not a higher relationship between post-estimates of improvement and actual improvement. The increased consistency from the second to the third trial in what little correlation there was between accuracy of estimate and over- and under-estimation in the three tasks also argues for general ''set'' in estimating within a given task.

Think- and May-Estimates

The position of the may-estimate on the estimate blank, coming as it does directly after the think-estimate, would naturally lead the subjects to base their may-estimates on their think-estimates. That such was the case in many instances is borne out by the statement of the subjects that their may-estimates were based on their think-estimates rather than on the original score. These estimates were made both by adding additional units to the think-estimates and by actually attempting to make a judicious estimate of possible improvement. The small difference found in the range of think- and may-estimates indicates that the rule of moderation held in making may-estimates as well as think-estimates. The may-estimates were, of course, higher than the think-estimates, but not markedly so. Incidentally only one subject reported that she took occasion to let herself go in the may-estimates. The difference in percent of accuracy between think- and may-estimates was not great. In cancellation, since the subjects were under-estimating improvement throughout, the may-estimates were more accurate than the think-estimates; in scribbling and knot tying the reverse was true.

On the whole, then, it seems probable that the process of making may-estimates consisted in following along the line of think-estimates. Any wide divergence in the latter estimate would simply be repeated in slightly exaggerated form in the former. In knot tying, this exaggeration resulted at times in a change from under-estimation to over-estimation since the original score was small.

Goal Estimates

It will be recalled that in the fourth trial the subjects were asked to set a goal of improvement over the original score instead of merely estimating how much they would gain or lose. The directions for cancellation that follow give the exact wording used:

"I want you to assume the attitude that you will improve on your first score. Then set a goal in actual units of improvement, *i. e.*, in additional series of numbers over your original score correctly examined. You may refer to your original score in setting your goal. Your goal may be as large or as small as you wish. Record the goal after the statement, 'Set as my goal an improvement of..............................' "

These directions set up quite a different situation than that called forth by the think- and may-estimates. The assumption that improvement would be made, for one thing, ruled out estimates of loss. The very fact that a distinction was being made between setting a goal and making an estimate carried with it the implication that the goal was something to aim at rather than a mere estimate of probable improvement. That these implications were grasped by the subjects became evident in the quantitative results obtained.

The first noticeable difference between the goal and other estimates was the increased size of the goal in comparison with the size of the pre-think estimates in the two previous trials. Figure 4 brings out this difference clearly in the sharp rise of the estimate curve for the goal trial in all three tasks. The performance, however, was not appreciably increased by setting a definite goal. As a result of the increased size of the estimate, therefore, there occurred a decrease in the percent of under-estimation in cancellation, and an increase in percent of over-estimation in scribbling and knot tying. This decided increase in the size of the goal might have been due to factors other than those inherent in the estimating process. The fact that the goal was set in the fourth trial, for instance, might be sufficient to account for its increased size, since the estimates were increasing from trial to trial. Or the fact that the goal trial was the last trial might have influenced the subjects to become a little less cautious. The omission of loss scores would of course automatically raise the estimate score. However, estimates of loss occurred so infrequently during the experiment that this factor is of small importance.

Since the position of the goal in the experiment did seem to be a contributing factor in the situation, a supplementary experiment was performed in order to discover whether the size of the goal would still be larger than other estimates if it occurred in a different place in the experiment. Fifteen graduate students were utilized for this investigation, meeting in three groups of five. One of these groups was directed to set the goal in the second trial, the other two groups in the third trial. Table 27 presents the results of this experiment, with the results found when the goal was placed in the fourth trial.

Before examining these results let us consider for a moment the significance of the different locations of the goal. In placing the goal in the second trial we are asking our subjects to set a goal before they have any experience of improvement. At the same time, they come to the task unprejudiced by previous estimates. Setting the goal in the third trial gives the subjects the benefit of a repetition of the task and of two experiences with estimating—before and after the second trial. In the fourth trial the goal is set after two repetitions of performance and four experiences in estimating. In addition it has the advantage, or disadvantage, of being the last trial.

On examining the results of the experiment we find that the setting of the goal raised the estimate no matter what its place in the experiment, for when the goal was placed in the second trial it was larger than both pre- and post-think estimates in the third trial in all three tasks. When the goal was placed in the third trial

TABLE 27

THE EFFECT OF GOAL POSITION ON THINK ESTIMATES

	Goal in Trial II *5 cases*			*Goal in Trial III* *10 cases*			*Goal in Trial IV* *406 cases*		
Trial	*Pre. Est.* %	*Post. Est.* %	*Imp.* %	*Pre. Est.* %	*Post. Est.* %	*Imp.* %	*Pre. Est.* %	*Post. Est.* %	*Imp.* %
				CANCELLATION					
II	*21	12	13	08	02	28	11	14	28
III	18	19	23	*16	05	40	15	16	34
IV	20	08	27	05	08	43	*32	20	43
				SCRIBBLING					
II	*29	15	01	09	12	07	17	15	12
III	15	16	08	*22	11	14	19	16	09
IV	33	29	15	12	11	06	*32	22	12
				KNOT TYING					
II	*36	21	25	07	05	09	18	12	18
III	32	05	21	*29	−07	06	19	11	20
IV	36	45	29	11	11	16	*38	21	26

*Goal.

it was larger than both pre- and post-estimates in either the second or fourth trial. And, as we have already seen, when it was placed in the fourth trial it was also larger than both the pre- and post-estimates in the two preceding trials. The difference in the size of the goal as compared with other estimates is therefore due in part to the nature of a goal. The subjects in other words set a goal to aim at, in distinction from an amount which they believe they will probably attain.

Since the goal in scribbling and knot tying was considerably greater than actual improvement, we conclude that setting the goal does not necessarily bring about a proportional increase in improvement. Actual improvement, however, was found (Table 19) to correlate positively with the size of the goal to a greater extent than with the other pre-estimates. This may indicate an influence of the goal upon actual improvement.

ESTIMATES OF LOSS

In our treatment of quantitative data we have made no distinction up to this point between estimates of gain and loss, other than considering the gain estimate a plus value and the loss estimate a minus value. As can be judged from the fact that the average estimates for all groups carried a positive sign, in all three tasks, the cases of estimated loss were few. In cancellation they represented only 4% of the total estimates, in scribbling 4%, and in knot tying 12%. These estimates, therefore, are relatively unimportant.

SAME-ESTIMATES

Estimates that the score would remain the same, or that the score had remained the same, were also few in number. In cancellation same-estimates occurred in only 6% of the possible cases, in scribbling 7%, and in knot tying 11%.

The same-estimates were made by subjects whose original scores were above the average for the whole group, and whose percent of improvement was therefore somewhat lower than the average, as brought out in the comparison below:

	Av. Original Score			*Av. Percent Improvement*		
	Can.	*Scrib.*	*Knots*	*Can.*	*Scrib.*	*Knots*
Whole Group	28.39	29.60	10.88	30	11	21
Same Group	28.88	31.27	11.68	26	06	16

Though those who estimated "same" made less improvement than the average for the whole group, they did improve, so that the estimate of "same" was in general an under-estimation.

It will be recalled that the directions for recording same-estimates read: "Those who checked 'Score will remain the same,' will fill in the two statements expressing possibility, 'May possibly gain as much as' and 'May possibly lose as much as.' " The purpose of asking for these two may-estimates was to see whether subjects when required to express an opinion as to possible gain or loss, even when they felt that the score would be the same, would estimate in favor of gain or loss. While the average may-gain estimates were somewhat larger than the may-lose, the difference between the two was only 1% in cancellation, 2% in scribbling and 1% in knot tying. These may-estimates were thus nearly consistent with the think-estimate of "same."

DEPARTMENT OF PSYCHOLOGY
LIBRARY
UNIVERSITY OF PORTLAND

VIII. SUMMARY AND CONCLUSIONS

Nature of the Problem

This investigation sets forth a procedure for studying an individual's ability to estimate his own improvement in a given task, and applies this procedure to a study of self-estimates in three simple tasks dependent largely on speed of perception and motor skill or dexterity, namely, cancellation, scribbling and knot tying.

The general plan of procedure was to have a subject perform a task, score his first performance, and repeat the task as many times as desired, estimating before and after each repetition the amount of improvement anticipated or actually made. The objective data obtained were supplemented by answers to a series of questions pertaining to the subject's method of estimating.

In the experiment described in this study, the three tasks were repeated three times in rotation, making four trials in all. Estimates before and after a trial were expressed in terms of improvement over the score in the first trial, the *original score,* which the subjects had before them. Two degrees of estimate were used: what the subjects *thought* they would gain or lose (or had gained or lost), and what they might *possibly* gain or lose (or had gained or lost). Freedom was also allowed to estimate that the score would remain, or had remained the same. The setting of a goal of improvement was introduced in the fourth trial.

The main experiment was performed by seven groups, totalling 406 adult subjects. Of these 346 were college students, 42 were high school students, and 18 were members of an executive training class in a retail store. Two supplementary groups were utilized later to secure introspective data on the estimating process and to examine the effect on the size of the goal of its position in the experiment.

Evaluation of the Method of Procedure

1. The general plan of procedure proved satisfactory in that it secured a variety of pertinent information in a short space of time, and was readily comprehended by the subjects.

2. The procedure could be simplified by omitting the "may-estimate" and the "same-estimates," since, on the basis of the present study, little additional information on the estimating process can be expected from these responses.

3. The objective character of the data obtained by the use of

this method makes comparison between estimates in different tasks and between different types of estimates simple. The introspective data make a valuable supplement in that they provide an insight into what is actually occurring during the estimating process.

4. It is felt that the use of an actual standard of reference rather than a guessed standard is an improvement over the usual technique employed in securing judgments, since it makes comparison between individual estimates possible.

5. The procedure could be utilized in estimating performance rather than improvement, by a slight change in wording.

6. The procedure would also lend itself to an intensive study of individual estimating.

7. By the introduction of slight variations in procedure, additional problems could be readily investigated, such as:

- *a.* The effect of practice on the estimating process, the results of each trial being known to the subject.
- *b.* A comparison of estimates expressed in time taken to complete a task with estimates of amount of work accomplished in a given time.
- *c.* The effect of estimates on improvement in a task, using a control group not estimating.
- *d.* The ability to estimate the improvement of others.
- *e.* Comparison of estimates made in performance under competitive and non-competitive conditions.

8. In order to get additional insight into the attitude of an individual towards his work this procedure could be applied to the work on the job. For example:

- *a.* Students could be requested to estimate their improvement (or performance) in daily work or on true-false tests, over a period of time. By analyzing the basis for their estimates some of the difficulties in their work might come to light.
- *b.* Workers could be asked to estimate production in terms of number or amount of sales they can make per day, or number of bills posted, etc. The reasons for failure to come up to the estimates made might afford some valuable hints on methods used.

RESULTS OF THE EXPERIMENT

Think-Estimates

Pre-Estimates.

1. There was a general tendency to estimate moderate improvement—an average estimate of about 17% of the original score for

the second and third trials (range 7% to 26%), and 35% for the goal (range 28% to 49%), differing with different groups and with the tasks. There were relatively few cases of estimated loss or that the score would remain the same.

2. There was a general tendency to raise estimates in successive trials.

3. Setting a goal increased the size of the estimate no matter in which trial in the experiment it was placed, but it did not necessarily increase the accuracy of estimates.

4. The average error for the pre-estimates in the second and third trials was approximately 20% of the original score, average error for the goal 22%. This indicates an accuracy which, while not high in comparison with such psychophysical judgments as those of intensity of light, length of line, etc., is much better than could be expected from mere guessing.

While there was a positive correlation between pre-estimates of improvement and actual improvement, the coefficients were not sufficiently high to indicate an individual's improvement from his estimate of improvement before trial.

5. There was no universal tendency to over- or under-estimate improvement. Under-estimation and over-estimation occurred in all three tasks, although in cancellation under-estimation predominated.

Over- or under-estimation was not an independent fact in this experiment but depended on performance as well as on estimates. Where actual improvement was small, over-estimation was the rule (as in scribbling); where improvement was rapid, the reverse was true (as in cancellation).

Estimates of improvement made before the subject had had an opportunity to try the task a second time showed less over- or under-estimation than pre-estimates after the subject had had experience with improvement.

Post-Estimates.

1. Post-estimates were more accurate than pre-estimates, but the advantage of the post-estimates was not very great in any one of the tasks.

2. There was a tendency in the post-estimates to correct or compensate for the under- or over-estimation of the pre-estimates.

3. There were rather higher correlations between post-estimates and performance than between pre-estimates and performance, but the coefficients in either case were not sufficiently high to be diagnostic of performance.

4. There was a fairly high positive correlation (coefficients well above 40) between pre- and post-estimates in the same trial, and between post-estimates in one trial and pre-estimates in the following trial.

May-Estimates

1. The may-estimates were not markedly higher or less accurate than the think-estimates.

The Three Tasks Compared

1. Cancellation showed the greatest possibility of improvement of the three tasks. However, this task was felt by the subjects to be difficult. The fact that the procedure was pretty well routinized tended to obscure variations in performance and to conceal the presence of improvement. Also fewer cues on which the subject might base his estimates were afforded than in the other two tasks. As a result under-estimation was found to be general in this task.

2. Scribbling, which involves a fairly automatic type of response without much skill, was the easiest for the subjects. It actually did not permit of much improvement, however, since as in tapping, the subjects tended to reach their physiological limit quickly. The presence of changes in methods of procedure during the progress of the experiment tended to raise the estimates of improvement in this task. While specific cues were present, they were of less definite help than those afforded in knot tying. Furthermore, these cues were sometimes misleading, as in the case of confusion between excess energy and increased output.

3. Knot tying called for quite a bit of skill in manipulation, and while it did permit of improvement, irregularities of performance, frequently of emotional origin, entered in. The fact that changes in method in this task afforded quite definite cues for estimating improvement accounted for the lack of any definite tendency towards over- or under-estimation in knot tying.

4. Knot tying stands out as the type of performance more nearly allied to everyday experience. There is room for a certain amount of improvement in this task, complicated, to be sure, by variations in performance, and this improvement is more or less self-evident. In scribbling improvement is restricted but seems possible; in cancellation improvement is easily possible but seems limited. As a result we get an even balance of over- and under-estimation in the first, over-estimation in the second, and under-estimation in the third.

Individual Consistency in Estimating

1. Some tendency towards "high" and "low" estimates was shown—some consistency in this respect, in the positive correlations found between size of estimates in the different tasks. The fact that the coefficients were well under 40 suggests that the difference in the nature of the tasks performed tended to counteract any marked tendency in this direction.

2. There was practically no individual consistency found in accuracy of estimates in the three tasks. The subjects were not consistently "good" or "poor" estimators regardless of the task performed.

3. The fact that there was even less consistency found between over- and under-estimation in the different tasks than between size of estimates leads us to believe that over- or under-estimation was not an independent fact, or an individual behavior trait in this experiment. Rather, it was a by-product of a combination of performance and estimation.

4. The small amount of individual consistency found between size of estimates and errors of estimates, in the different tasks, indicates that the estimating process was not a general behavior trait but was dependent on the task performed.

The Process of Estimating

1. First pre-estimates, those made before the subject had repeated the task, were based on a general belief in ability to improve with practice, plus a perception of improved methods of performance that could be adopted in scribbling and knot tying. Through scoring the first trial the subjects became aware of the type of scribbling that would produce the highest score. In knot tying the anticipated avoidance of errors made in the first trial, such as tying knots on top of each other or attempting to go too fast, was an important factor in making the first estimate.

2. In post-estimates the main cue utilized in cancellation was the position on the sheet that the subject had reached when time was called. In scribbling the character of the loops made (large or small, sharp or round), the general appearance of the page, distance between loops, number of columns made and tenseness of muscle, were given as indications of improvement, together with changes in method of procedure such as holding the paper differently, pulling the paper under the pencil, changing the arm position. Knot tying afforded the most definite cues, such as the number of double

knots tied, the smaller amount of space between the knots and the ''feel'' of the knots as they slipped through the fingers. Changes in procedure were also mentioned in this task, such as making looser and different types of knots, turning the string around and starting from both ends, holding the knot in place and pulling the string through, working clear of the table, removing finger ring. That actually counting the knots made, against instructions (or lines covered in cancellation, or scribbles made), was not resorted to frequently is evident from the fact that there was so little difference between the accuracy of the pre- and post-estimates. If counting had been the rule, the post-estimates would have been nearly 100% accurate.

3. Perception of speed of performance of separate acts, and smoothness of performance, had preference as criteria for estimating improvement in the three tasks. In knot tying, change of method of performance was more important as a criterion for estimates than in the other two tasks, while the direct estimate of improvement regardless of what happened during the trial held a higher place in cancellation.

4. The subjects reported that the size of the original score was used as the basis for pre-estimation. This statement was supported by the fact that the percent estimates of those with the highest original scores were lower than the percent estimates of those with the lowest original scores. That the size of the original score was not taken into account sufficiently to avoid error, however, was evident from the fact that the subjects having the highest original scores over-estimated more and under-estimated less than those with the lowest scores.

5. Within a given task, the subjects showed a tendency to establish a certain set in estimating improvement as brought out by the similarity between think- and may-estimates, and by the positive correlations found between pre- and post-estimates in the same trial, and between post-estimates in one trial and pre-estimates in the following trial. This dependence of the different estimates on each other was recognized by the subjects in the case of the pre- and post-estimates in the same trial, but not altogether in the case of post-estimates in one trial and pre-estimates in the following trial, since they reported that the post-estimates in one trial were frequently made on the pre-estimate in that trial, but that their estimates were made fairly independently for each trial.

6. The fact that the first pre-estimates were on the whole more accurate than subsequent estimates suggests that the set in estimat-

ing became more firmly established as the experiment progressed. This is readily understood when we realize that the subjects had no way of checking up on the accuracy of their estimates, or on the amount of improvement made, since they were not given their scores after the first trial.

7. May-estimates were based on the think-estimates rather than on the original score, and were made by adding units to the think-estimates or by actually attempting to make a cautious estimate of possible improvement.

8. From the results of the experiment we conclude that the process of estimation was dependent to a greater or less extent on the following factors:

The apparent possibility of improvement in the given task
The size of the original score
Changes in method of performance during the task
The emotional attitude of the subject towards the task
The set of the estimator established at the beginning of the experiment
Familiarity with the task, and with the estimating procedure
The type of estimate to be made
The nature of the task to be performed
The individual making the estimate.

IX. BIBLIOGRAPHY

1. Allport, F. H., and Allport, G. W. Personality traits, their classification and measurement. *J. Abn. Psychol.,* 1921–22, 16, 1–40.
2. Brown, F. J. Knowledge of results as an incentive in school room practice. *J. Educ. Psychol.,* 1932, 23, 532–552.
3. Cattell, J. McK. The psychological researches of James McKeen Cattell. *Arch. Psychol.,* 1914, No. 14.
4. Farmer, E. Concerning subjective judgments of difficulty. *Brit. J. Psychol.,* 1928, 18, 438–442.
5. Filter, R. O. Experimental study of character traits. *J. Appl. Psychol.,* 1921, 5, 297–317.
6. Filter, R. O. Estimates of amount of work one can do. *J. Appl. Psychol.,* 1927, 11, 58–67.
7. Fisher, R. A. Statistical methods for research workers. Edinburgh, Oliver, 1932.
8. Fullerton, G. S., and Cattell, J. McK. On the perception of small differences. Univ. Penn. Press, 1892.
9. Gates, G. S. Effect of encouragement and discouragement upon performance. *J. Educ. Psychol.,* 1923, 14, 21–26.
10. Gates, G. S. Effect of audience upon performance. *J. Abn. Psychol.,* 1923–24, 18, 334–344.
11. Gilliland, A. R. A revision and some results with the Moore-Gilliland aggressiveness test. *J. Appl. Psychol.,* 1926, 10, 143–150.
12. Granich, L. A technique for experimentation on guessing in objective tests. *J. Educ. Psychol.,* 1931, 22, 145–156.
13. Greene, E. B. Achievement and confidence on true-false tests of college students. *J. Abn. Psychol.,* 1929, 23, 467–478.
14. Henmon, V. A. C. The relation of the time of judgment to its accuracy. *Psychol. Rev.,* 1911, 18, 186–201.
15. Hoffman, G. J. An experiment in self-estimation. *J. Abn. Psychol.,* 1923–24, 18, 43–49.
16. Hollingworth, H. Experimental studies in judgment. *Arch. Psychol.,* 1915, No. 29.
17. Langdon, J. N. A note on the repetition of a simple motor test. *Brit. J. Psychol.,* 1931, 22, 55–61.
18. Lund, F. H. The psychology of belief. *J. Abn. Psychol.,* 1925–26, 20, 63–81; 174–195.
19. Morgan, J. B. The measurement of persistence. *J. Appl. Psychol.,* 1926, 10, 180–187.
20. Peterson, J. Johnson's measurement of rate of improvement under practice. *J. Educ. Psychol.,* 1924, 15, 271–275.
21. Travis, L. Effect of a small audience upon eye-hand coördination. *J. Abn. Psychol.,* 1925–26, 20, 142–146.
22. Trow, W. C. Psychology of confidence. *Arch. Psychol.,* 1924, No. 67.
23. Trowbridge, M. H., and Cason, H. An experimental study of Thorndike's theory of learning. *J. Gen. Psychol.,* 1932, 7, 245–260.
24. Turney, A. H. Study of judgments. *J. Appl. Psychol.,* 1931, 15, 259–272.
25. Whittemore, I. Influence of competition on performance. *J. Abn. Psychol.,* 1924–25, 19, 236–253.
26. Whittemore, I. The competitive consciousness. *J. Abn. Psychol.,* 1925, 20, 17–33.
27. Wyatt, S. Maximum capacity and average achievement. *Brit. J. Psychol.,* 1929–30, 20, 251–260.

R.W.B. JACKSON LIBRARY
3 0005 03037627 4

150.8
A673
no.163
Archives of Psychology

150.8
A673
no.163
Archives of Psychology

ARCHIVES OF PSYCHOLOGY

List of numbers, continued from inside front cover

88. Mental Hygiene Inventory: S. D. House. $1.50.
89. Mental Set and Shift: A. T. Jersild. $1.25.
90. Experimental Investigation of Rest Pauses: C. W. Manzer. $1.25.
91. Routine and Varying Practice as Preparation for Adjustment to a New Situation: L. W. Crafts. $1.00.
93. Speed and Other Factors in "Racial" Difference: O. Klineberg. $1.50.
94. Relation of Reaction Time to Intelligence, Memory, and Learning: V. W. Lemmon. 80c.
95. Is the Latent Time in the Achilles Tendon Reflex a Criterion of Speed in Mental Reactions? G. H. Rounds. $1.25.
96. Predictive Value of Tests of Emotional Stability Applied to College Freshmen: E. G. Flemming. $1.00.
97. Vocabulary Information Test: A. L. Weeks. $1.00.
98. Effect of Temporal Arrangements of Practice on the Mastery of an Animal Maze: S. A. Cook. 80c.
99. Recognition Time as a Measure of Confidence: G. H. Seward. $1.00.
100. Precision and Accuracy: G. W. Hartman. 80c.
101. Group Test of Home Environment: E. M. Burdick. $1.50.
102. Effect of Material on Formal Syllogistic Reasoning: M. C. Wilkins. $1.25.
103. Effect of Incentives on Accuracy of Discrimination: H. C. Hamilton. $1.25.
104. Validity of Norms with Special Reference to Urban and Rural Groups: M. E. Shimberg. $1.25.
105. Blood Pressure Changes in Deception: M. N. Chappell. 80c.
106. Experimental Comparison of Psycho-physical Methods: W. N. Kellogg. $1.25.
107. Measurement of Verbal and Numerical Abilities: M. M. R. Schneck. $1.00.
108. Perseverative Tendency in Pre-School Children. A Study in Personality: H. M. Cushing. $1.00.
109. Preliminary Study of the Effect of Training in Junior High School Shop Courses: L. D. Anderson. 80c.
110. Music Appreciation: M. J. Adler. $1.50.
111. Motivation in Fashion: E. B. Hurlock. $1.00.
112. Equality Judgments in Psychophysics: W. N. Kellogg. $1.00.
113. Illusions in the Perception of Short Time Intervals: N. Israeli. 80c.
114. Further Studies of the Reading-Recitation Process in Learning: Skaggs, Grossman, Krueger & Krueger. 80c.
115. Factors Affecting the Galvanic Reflex: R. C. Davis. $1.00.
116. Infant's Feeding Reactions During the First Six Months: R. Ripin. 80c.
117. Measurement of Mental Deterioration: H. Babcock. $1.25.
118. Phenomenon of Postural Persistence: L. S. Selling. $1.00.
119. American Council on Education Rating Scale: F. F. Bradshaw. $1.00.
120. Group Factor in Immediate Memory: A. Anastasi. $1.00
121. Individual Differences in the Sense of Humor and Temperamental Differences: P. Kambouropoulou. $1.00.
122. Suggestibility in Normal and Hypnotic States: G. W. Williams. $1.00.
123. Analytical Study of the Conditioned Knee-Jerk: G. R. Wendt. $1.25.
124. Race Differences in the Organization of Numerical and Verbal Abilities: J. W. Dunlap. $1.25.
125. Errors of Measurement and Correlation: E. E. Cureton. $1.25.
126. Experience Factors, Test Scores and Efficiency of Women Office Workers: N. Bird. $1.00.
127. Delayed Reactions of Infants: C. N. Allen. 80c.
128. Factors Measured by the Thorndike Intelligence Examination for H. S.: J. G. Peatman. $1.00.
129. Educational Success and Failure in Supernormal Children: J. Regensburg. $1.75.
130. Effect of Practice on Visual Perception of Form: J. P. Seward. $1.00.
131. Relation to College Grades of Some Factors other than Intelligence: D. Harris. 80c.
132. Study of Psychological Differences Between "Racial" and National Groups in Europe: O. Klineberg. $1.00.
133. Emotional Differences of Delinquent and Non-Delinquent Girls of Normal Intelligence: A. Courthial. $1.25.
134. Learning and Retention of Pleasant and Unpleasant Activities: H. Cason. $1.25.
135. Investigation of Brightness Constancy: R. B. MacLeod. $1.25.
136. The Rorschach Test Applied to Feeble-Minded Group: S. J. Beck. $1.00.
137. Retention after Intervals of Sleep and of Waking: E. B. van Ormer. $1.00.
138. Stimulus Temperature and Thermal Sensation: F. Heiser. $1.00.
139. Energy Cost Measurements on Curve of Work: H. J. Schubert. $1.00.
140. Technique for the Measurements of Attitudes: R. Likert. 80c.
141. Speed Factor in Mental Tests: P. H. DuBois. 80c.
142. Further Studies on the Memory Factor: A. Anastasi. $1.00.
143. An Experimental Study on Variability of Learning: S. E. Asch. $1.00.
144. Development of Inventory for Measurement of Inferiority Feelings at H. S. Level: R. B. Smith. $1.50.
145. The Psychological Effects of Oxygen Deprivation: R. A. McFarland. $1.50.
146. Relation of Subliminal to Supraliminal Learning: O. A. Simley. 80c.
147. Effects of Noise upon Certain Psychological Processes: F. L. Harmon. $1.25.
148. Conditioned Responses in Children: G. S. Razran. $1.50.
149. Resemblance of Parents and Children in General Intelligence: M. C. Outhit. $1.00.
150. Influence of Oral Propaganda Material . . . : K-C. Chen. 80c.
151. Negative or Withdrawal Attitude: H. Pallister. 80c.
152. Judgment in Absolute Units as a Psychophysical Method: J. Bressler. $1.00.
153. Visual Illusions in the Chick: C. N. Winslow. $1.25.
154. Measuring Teaching Efficiency Among College Instructors: G. W. Hartmann. 80c.
155. Psychogalvanic Responses in Arithmetical Work: R. Sears. $1.00.
156. Group Factors in Mental Tests: M. Smith. $1.00.
157. . . . Brightness Discrimination Habit in the Chick: A. B. Wood. 80c.
158. Accuracy of perception of visual musical stimuli: A. Roe. $1.00.
159. Effect of Practice Upon Individual Differences: R. E. Perl. $1.00.
160. Antagonistic Muscle Action During the Initiatory Stages of Voluntary Effort: D. J. Wilson. $1.00.
161. Verbal, Numerical and Spatial Abilities of Young Children: B. Schiller. $1.00.
162. Organization of Memory in Young Children: A. I. Bryan. $1.00.
163. Self-Estimates of Improvement in Repeated Tasks: N. Kneeland. $1.25.

DIRECTORY OF AMERICAN PSYCHOLOGICAL PERIODICALS

American Journal of Psychology—Ithaca, N. Y.; Cornell University.
Subscription $6.50. 624 pages annually. Edited by M. L. Washburn, K. M. Dallenbach, Madison Bentley, and E. G. Boring. Quarterly. General and experimental psychology. Founded 1887.

Journal of Genetic Psychology—Worcester, Mass.; Clark University Press.
Subscription $14.00 per year; $7.00 per volume. 1000 pages annually. (2 volumes.) Edited by Carl Murchison. Quarterly. Child behavior, animal behavior, and comparative psychology. Founded 1891.

Psychological Review—Princeton, N. J.; Psychological Review Company.
Subscription $5.50. 540 pages annually. Edited by Howard C. Warren. Bi-monthly. General psychology. Founded 1894.

Psychological Monographs—Princeton, N. J.; Psychological Review Company.
Subscription $6.00 per volume. 500 pages. Edited by Herbert S. Langfeld. Published without fixed dates, each issue one or more researches. Founded 1895.

Psychological Index—Princeton, N. J.; Psychological Review Company.
Subscription $4.00. 300–400 pages. Edited by Walter S. Hunter and R. R. Willoughby. An annual bibliography of psychological literature. Founded 1895.

Psychological Bulletin—Princeton, N. J.; Psychological Review Company.
Subscription $6.00. 720 pages annually. Edited by Edward S. Robinson. Monthly (10 numbers). Psychological literature. Founded 1904.

Archives of Psychology—New York, N. Y.; Columbia University.
Subscriptions $6.00. 500 pages per volume. Edited by R. S. Woodworth. Without fixed dates, each number a single experimental study. Founded 1906.

Journal of Abnormal and Social Psychology—Eno Hall, Princeton, N. J.; American Psychological Association.
Subscription $5.00. 448 pages annually. Edited by Henry T. Moore. Quarterly. Abnormal and social. Founded 1906.

Psychological Clinic—Philadelphia, Pa.; Psychological Clinic Press.
Subscription $3.00. 288 pages. Edited by Lightner Witmer. Without fixed dates. (9 numbers). Orthogenics, psychology, hygiene. Founded 1907.

Psychoanalytic Review—Washington, D. C.; 3617 10th St., N. W.
Subscription $6.00. 500 pages annually. Edited by W. A. White and S. E. Jeliffe. Quarterly. Psychoanalysis. Founded 1913.

Journal of Experimental Psychology—Princeton, N. J.
Psychological Review Company. 700 pages annually. Experimetal. Subscription $7.00. Bimonthly. Edited by S. W. Fernberger. Founded 1916.

Journal of Applied Psychology—Athens, Ohio.
Subscription $5.50. 400 pages annually. Edited by James P. Porter. Bi-monthly. Founded 1917.

Journal of Comparative Psychology—Baltimore, Md.; Williams & Wilkins Company.
Subscription $5.00 per volume of 450 pages. Three volumes every two years. Edited by Knight Dunlap and Robert M. Yerkes. Founded 1921.

Comparative Psychology Monographs—Baltimore, Md.; The Johns Hopkins Press.
Subscription $5.00. 400 pages per volume. Knight Dunlap, Managing Editor. Published without fixed dates, each number a single research. Founded 1922.

Genetic Psychology Monographs—Worcester, Mass.; Clark University Press.
Subscription $14.00 per year; $7.00 per volume. 1000 pages annually. (2 volumes.) Edited by Carl Murchison. Monthly. Each number one complete research. Child behavior, animal behavior, and comparative psychology. Founded 1925.

Psychological Abstracts—Eno Hall, Princeton, N. J.; American Psychological Association.
Subscription $6.00. 700 pages annually. Edited by Walter S. Hunter and R. R. Willoughby. Monthly. Abstracts of psychological literature. Founded 1927.

Journal of General Psychology—Worcester, Mass.; Clark University Press.
Subscription $14.00 per year; $7.00 per volume. 1000 pages annually. (2 volumes.) Edited by Carl Murchison. Quarterly. Experimental, theoretical, clinical, and historical psychology. Founded 1927.

Journal of Social Psychology—Worcester, Mass.; Clark University Press.
Subscription $7.00. 500 pages annually. Edited by John Dewey and Carl Murchison. Quarterly. Political, racial, and differential psychology. Founded 1929.